

02-07-00
IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

ATTY DOCKET NO.: 5470-130DV

DATE: February 3, 2000

UTILITY PATENT APPLICATION TRANSMITTAL LETTER
AND FEE TRANSMITTAL FORM (37 CFR 1.53(b))

BOX PATENT APPLICATION

Assistant Commissioner for Patents

Washington, DC 20231

Sir:

Transmitted herewith for filing under 37 CFR 1.53(b) is:

- ☒ a patent application
☐ a Continuation ☒ a Divisional ☐ a Continuation-in-Part (CIP)
of prior application no.: 07/182,646; filed 15 April 1988.
☐ A Small Entity Statement(s) was filed in the prior application; Status still proper and desired.

Inventor(s) or Application Identifier:

Frank S. French, Chapel Hill, North Carolina;
Elizabeth M. Wilson, Chapel Hill, North Carolina
David R. Joseph, Chapel Hill, North Carolina
Dennis B. Lubahn, Columbia, Missouri

Entitled: **ANDROGEN RECEPTOR PROTEINS, RECOMBINANT DNA MOLECULES CODING
FOR SUCH, AND USE OF SUCH COMPOSITIONS**

Enclosed are:

1. ☒ Application Transmittal Letter and Fee Transmittal Form (*A duplicate is enclosed for fee processing*)
2. ☒ 26 pages of Specification (including 6 claims)
3. ☒ 22 sheets of Drawings (35 USC 113)
4. ☐ Oath or Declaration
 - a. ☐ newly executed (*original or copy*)
 - b. ☐ copy from prior application (37 CFR 1.63(d) (*for continuation/divisional*) [Note Box 5 Below]
 - c. ☐ DELETION OF INVENTOR(S) (*Signed statement deleting inventor(s) named in the prior application*)
5. ☐ Incorporation By Reference (*useable if box 4b is checked*)

The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein.
6. ☐ Microfiche Computer Program (*Appendix*)
7. ☒ Assignment papers (*cover sheet(s) and document(s)*) **Filed in parent application**
8. ☐ Small Entity Statement(s)
9. ☒ Information Disclosure Statement, PTO-1449, and 0 of 25 references cited
10. ☒ Preliminary Amendment (*Please enter all claim amendments prior to calculating the filing fee.*)

11. ☐ English Translation Document
 12. ☐ Certified Copy of
 13. ☐ Sequence Listing/ Sequence Listing Diskette
 a. ☐ computer readable copy
 b. ☐ paper copy
 c. ☐ statement in support
 14. ☐ An Associate Power of Attorney
 15. ☒ Return Receipt Postcard (MPEP 503) *(Should be specifically itemized)*
 16. ☒ Other: Copies of: Decision to Revive Patent Application, Terminal Disclaimer,
 The fee has been calculated as shown below:

	Column 1 No. Filed	Column 2 No. Extra	Small Entity Rate Fee	Large Entity Rate Fee
BASIC FEE			\$345.00	\$690.00
TOTAL CLAIMS	8 - 20 =	0	x 09 = \$	x 18 = \$
INDEP CLAIMS	2 - 3 =	0	x 39 = \$	x 78 = \$
<input type="checkbox"/> MULTIPLE Dependent Claims Presented			+ 130 = \$	+ 260 = \$
<i>If the difference in Col. 1 is less than zero, Enter "0" in Col. 2</i>			Total \$	Total \$690.00

- ☒ A check in the amount of \$690.00 to cover the filing fee is enclosed.
- ☐ A check in the amount of \$ is enclosed to cover the filing fee, PLUS the Assignment Recordation fee (\$40.00).
- ☒ The Commissioner is hereby authorized to charge payment of the following fees associated with this communication or credit any overpayment to Deposit Account No. 50-0220.
- ☒ Any additional filing fees required under 37 CFR 1.16.
- ☒ Any patent application processing fees under 37 CFR 1.17.

Respectfully submitted,

Karen Magri

Karen A. Magri
 Registration No. 41,965

Correspondence Address:
 USPTO Customer Number: **20792**
 Myers Bigel Sibley & Sajovec
 Post Office Box 37428
 Raleigh, North Carolina 27627
 Telephone (919) 854-1400
 Facsimile (919) 854-1401

CERTIFICATE OF EXPRESS MAILING

Express Mail Label No. EL533607859US
 Date of Deposit: February 3, 2000

I hereby certify that this correspondence is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to Box Patent Application, Assistant Commissioner For Patents, Washington, DC 20231.

Marjorie J. Pfeiffer
 Marjorie J. Pfeiffer
 Date of Signature: February 3, 2000

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re: Application of Frank S. French et al.
Serial No.: To be assigned
Filed: Concurrently herewith
For: *ANDROGEN RECEPTOR PROTEINS
RECOMBINANT DNA MOLECULES
AND CODING FOR SUCH USE OF
SUCH COMPOSITIONS*

Date : February 3, 2000

BOX PATENT APPLICATION
Assistant Commissioner for Patents
Washington, DC 20231

PRELIMINARY AMENDMENT

Sir:

Applicants respectfully request entry of the following amendment in the above-referenced application.

In the Claims:

Please cancel Claims 1-4 for the purposes of rewriting. Please cancel Claims 5 and 6, which were prosecuted in the parent application.

7. An isolated and purified DNA sequence encoding human androgen receptor.

8. The isolated and purified DNA sequence according to claim 7, said receptor having the amino acid sequence set forth in Figure 5.

9. The isolated and purified DNA sequence encoding human androgen receptor and as set forth in Figure 5.

10. The isolated and purified DNA sequence according to claim 9, said DNA sequence having the nucleotide sequence as set forth in Figure 5.

11. A human androgen receptor protein encoded by the DNA according to any one of claim 7 to 10.

12. A prokaryotic or eukaryotic host cell transformed or transfected with a DNA sequence according to any one of claims 7 to 10.

13. A viral or circular DNA plasmid comprising a DNA sequence according to any one of claims 7 to 10.

14. The viral or circular DNA plasmid according to claim 13 further comprising an expression control sequence operatively associated with said DNA sequence.

In the Specification:

Please make the following amendments to the specification.

On page 1 of the application, after the title, add the following:

--Related Applications

This application is a divisional of United States Application Serial No. 07/182,646, filed on April 15, 1988, which is hereby incorporated by reference in its entirety.--

REMARKS

The present amendment is submitted to complete the record, to present additional claims for substantive examination, and to provide the basis for an interference with U.S. Patent No. 5,614,620 to Liao et al., issued March 25, 1997.

In re: Application of Frank S. French et al.
Serial No.: To be assigned
Filed: Concurrently herewith
Page 3 of 3

Pursuant to 37 CFR 1.607(c), it is noted that claims 7-10 and 12-14 presented above correspond to Claims 1 and 6-9 in U.S. Patent No. 5,614,620 to Liao et al. Other claims previously of record, particularly Claim 2 and Claims 7-14 in the parent application, are substantially the same as those submitted in Liao. Claims 2 and 7-14 were of record in the present case prior to or within one year of the issuance of the Liao patent. It is noted that the parent application was filed within three months subsequent to the filing of the Liao patent and, accordingly, falls within the provisions of 37 C.F.R. § 1.608(a).

Applicants respectfully submit that the present application is in condition for substantive examination, which action is respectfully requested.

Respectfully submitted,

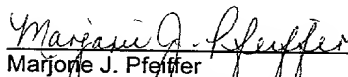


Karen A. Magri
Registration No. 41,965

Myers Bigel Sibley & Sajovec
PO Box 37428
Raleigh NC 27627
Telephone (919) 854-1400
Facsimile (919) 854-1401

"Express Mail" mailing label number EL533607857US
Date of Deposit: February 3, 2000

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to Box PATENT APPLICATION, Assistant Commissioner of Patents, Washington, DC 20231.



Marjorie J. Pfeffer
Date of Signature: February 3, 2000

ANDROGEN RECEPTOR PROTEINS, RECOMBINANT DNA MOLECULES CODING
FOR SUCH, AND USE OF SUCH COMPOSITIONS

5 This invention was made in the course of research supported in part by grants from the National Institutes of Health (NIH HD 16910, HD 04466, and HD 18968).

TECHNICAL FIELD OF THE INVENTION

10 This invention relates to recombinant DNA molecules and their expression products. More specifically this invention relates to recombinant DNA molecules coding for androgen receptor protein, androgen receptor protein, and use of the DNA molecules and protein in investigatory, diagnostic and therapeutic applications.

15 BACKGROUND OF THE INVENTION

 The naturally occurring androgenic hormones, testosterone and its 5 -reduced metabolite, dihydrotestosterone, are synthesized by the Leydig cells of the testes and circulate throughout the body where they diffuse into cells and bind to the androgen receptor protein ("AR"). Androgens, acting through their receptor, stimulate development of the male genitalia and accessory sex glands in the fetus, virilization and growth in the pubertal male, and maintenance of male virility and reproductive function in the adult. The

20 androgen receptor, together with other steroid hormone receptors constitute a family of trans-acting transcriptional regulatory proteins that control gene transcription through interactions with specific gene sequences.

25 When prostate cancer is found to be confined to the prostate gland, the treatment of choice is surgical removal. However, 50 to 80% of prostate cancer patients already have metastases at the time of diagnosis. Most of their tumors (70 to 80%) respond to the removal of androgen by castration or by suppression of luteinizing hormone secretion by the pituitary gland using a gonadotropin

30

releasing hormone analogue alone or in combination with an anti-androgen. The degree and duration of response to this treatment is highly variable (10% live < 6 months, 50% live < 3 years, and 10% live > 10 years.) Initially cancer cells regress without androgen stimulation, but ultimately the growth of androgen independent tumor cells continues (3b). At present it is not possible to predict on an individual basis which patient will respond to hormonal therapy and for how long. If poorly responsive patients could be identified early, they could be treated by alternative forms of therapy (e.g. chemotherapy) at an earlier stage when they might be more likely to respond.

Studies on androgen receptors in prostate cancer have suggested that a positive correlation may exist between the presence of androgen receptors in cancer cells and their dependence on androgenic hormone stimulation for growth. (An analogous situation exists in mammary carcinoma where there is a correlation between estrogen receptors and regression of the tumor in response to estrogen withdrawal). However, methodological problems in the measurement of androgen receptors have prevented the routine use of androgen receptor assays in the diagnostic evaluation of prostate cancer. Prior to our preparation of androgen receptor antibodies, all androgen receptor assays were based on the binding of [³H]-labeled androgen. These assays have been unreliable in human prostate cancer tissue because of the extreme lability of the androgen binding site and the presence of unlabeled androgen in the tissue. Endogenous androgen occupies the binding site on the receptor and dissociates very slowly (t 1/2 24-48 hr at 0C). A further problem is that biopsy samples are quite small, making it difficult to obtain sufficient tissue for [³H]-androgen binding assays. Moreover, prostate cancer is heterogenous with respect to cell types. Thus within a single biopsy sample there is likely to be an uneven distribution of cells containing androgen receptors.

Development of the male phenotype and maturation of male reproductive function are dependent on the interaction of androgenic

hormones with the androgen receptor protein and the subsequent function of the receptor as a trans-acting inducer of gene expression. It has become well established over the past twenty-five years that genetic defects of the androgen receptor result in a broad spectrum of developmental and functional abnormalities ranging from genetic males (46,XY) with female phenotype to phenotypically normal males with infertility. Isolation of the structural gene for the androgen receptor makes it possible to define the nature of these genomic defects in molecular terms. Analysis of the functional correlates of the genetic defects may lead to a better understanding of the regulation of androgen receptor gene expression and of the mechanism of androgen action in male sexual development and function.

The androgen insensitivity syndrome, known also as testicular feminization, is characterized by an inability to respond to androgen due to a defect in the androgen receptor, the protein that mediates the action of androgen within the cell. Androgen insensitivity is an inherited X-linked trait that occurs in both complete and incomplete forms. The complete form results in failure of male sex differentiation during embryogenesis and absence of virilization at puberty. The result is a 46,XY genetic male with testes and male internal ducts. The testes produce normal amounts of testosterone and Mullerian inhibiting substance. Consequently development of the uterus is inhibited as in the normal male. Because of the inability to respond to androgen, the external genitalia remain in the female phenotype with normal clitoris and labia. A small vagina develops from the urogenital sinus and ends in a blind pouch. At puberty feminization with breast development and female contours occur in response to testicular estrogen, however, there is no growth of sexual hair even though circulating testosterone concentrations are equal to or greater than levels in the normal male.

Incomplete forms of the androgen insensitivity syndrome include a spectrum of phenotypes resulting from varying degrees of

incomplete androgen responsiveness. At one extreme, individuals have mild enlargement of the clitoris and sparse pubic hair. The opposite extreme is characterized by more complete masculinization with varying degrees of hypospadias deformity but predominantly a male phenotype. It has been reported that some adult men with severe oligospermia or azoospermia who are otherwise normal, have defects in the androgen receptor. These may include as many as 10% of infertile males.

The genetic defect eliciting this range of abnormalities is thought to be a single biochemical event at the level of the gene for the androgen receptor. The androgen receptor is a high affinity androgen binding protein that mediates the effects of testosterone and dihydrotestosterone by functioning as a trans-acting inducer of gene expression. For proper male sexual development to occur, there is a requirement for androgen and its receptor at a critical time during embryogenesis and during puberty. The majority of individuals with the androgen insensitivity syndrome have a history of affected family members, although about a third are thought to represent new mutations of this X-linked disorder. The incidence ranges from 1 in 20,000 to 60,000 male births.

In studies of families with clinical evidence of the androgen insensitivity syndrome, four main categories were recognized that range from the most severe, complete absence of receptor binding activity in a genetic male with female phenotype, to qualitatively normal receptor in affected individuals. Second in severity are affected individuals with qualitatively abnormal androgen binding by receptor present in normal levels. Examples include the failure of sodium molybdate (a reagent often used in studies on steroid receptors) to stabilize the receptor of affected individuals when molybdate is known to stabilize the wild-type receptor. Lability of the receptor under conditions that normally cause transformation has also been reported. A third group expresses a decreased amount of receptor with wild-type in vitro binding characteristics. The final grouping contains those androgen insensitivity patients in

whom no abnormality in receptor is detected. In a recent study of this form of the syndrome, the androgen receptor was as capable of binding oligonucleotides as the wild-type receptor. Indeed, with the techniques available until only recently, it has been difficult
5 in certain cases to document an androgen receptor defect in affected individuals.

Experimental methods used in assessing receptor defects in the past have relied on the ability of receptor to bind androgen with high affinity. The limitation of this methodology is that it is not
10 possible to distinguish between the lack of expression of the receptor and loss of androgen binding activity. An example of how inadequate methodology complicates diagnosis is the absence of detectable receptor binding activity in patients who are partially virilized. It is theoretically possible for a mutation to occur
15 which allows the receptor with defective androgen binding activity to induce gene transcription. Biologically active truncated forms of the glucocorticoid receptor that lack steroid binding activity but retain the DNA binding domain have been demonstrated using genetically engineered mutants.

Purification of the androgen receptor has been difficult to accomplish due to its low concentration and high degree of instability. Reported attempts at purification using either
20 conventional methods of column chromatography or steroid-affinity chromatography have yielded insufficient amounts of receptor protein to allow even the preparation of monoclonal antibodies.
25

An early report on the partial purification of the androgen receptor was disclosed by Mainwaring et al. in "The use of DNA - cellulose chromatography and isoelectric focusing for the
30 characterization and partial purification of steroid-receptor complexes," Biochem J, 134, 113-127 (1973). They used DNA-cellulose chromatography and isoelectric focusing to isolate the receptor from rat ventral prostate and determined its physiochemical properties. This group was among the first to attempt the use of steroid affinity chromatography in conjunction with conventional

chromatography, using the affinity label
17B-bromoacetoxytestosterone in receptor purification (See
Mainwaring et al., "Use of the affinity label
17B-bromoacetoxytestosterone in the purification of androgen
5 receptor proteins," Perspectives in Steroid Receptor Research,
(1980)). Partial purification of androgen receptor has also been
attempted from other tissue sources, such as ram seminal vesicles
(See Foekens et al., Molecular Cellular Endocr, 23, 173-186 (1981)
and Foekens et al., "Purification of the androgen receptor of sheep
10 seminal vesicles," Biochem Biophys Res Comm, 104, 1279-1286
(1982)). The partially purified receptor displayed characteristics
of a proteolyzed receptor, but a purification of 2,000 fold was
reported with a recovery of 33% (See Foekens et al., "Purification
of the androgen receptor of sheep seminal vesicles," Biochem Biophys
15 Res Comm, 104, 1279-1286 (1982)). Later attempts at purification
continued to combine steroid affinity chromatography with
conventional techniques, reportedly achieving significant
purification, but recoveries too low for further analysis (See Chang
et al., "Purification and characterization of androgen receptor from
20 steer seminal vesicle," Biochemistry 21, 4102-4109 (1982), Chang et
al., "Purification and characterization of the androgen receptor
from rat ventral prostate," Biochemistry 22, 6170-6175 (1983) and
Chang et al., "Affinity labeling of the androgen receptor in rat
prostate cytosol with
25 17B-[(bromoacetyl)oxy]-5-alpha-androstan-3-one," Biochemistry 23,
2527-2533 (1984)). More recent studies examine the effectiveness of
a variety of immobilized androgens for their ability to bind the
androgen receptor (See De Larminat et al., "Synthesis and evaluation
of immobilized androgens for affinity chromatography in the
30 purification of nuclear androgen receptor," The Prostate 5, 123-140
(1984) and Bruchovsky et al., "Chemical demonstration of nuclear
androgen receptor following affinity chromatography with immobilized
ligands," The Prostate 10, 207-222 (1987)). Despite these efforts,
the receptor has not been purified to homogeneity and

the quantities of purified androgen receptor obtained have been insufficient for the production of antisera.

Clinical assays for the androgen receptor now include several methods. The most common is the binding of tritium-labeled hormone and measurement of binding using a charcoal adsorption assay. Either a natural androgen, such as dihydrotestosterone, or synthetic androgen, such as mibolerone or methyltrienolone (R1881), can be used. An advantage of the latter in human tissue is that it is not significantly metabolized and does not bind to the serum androgen binding protein, sex steroid binding globulin. A limitation of radioisotope labeling of receptor is interference caused by endogenous androgen. Although exchange assays for the androgen receptor have been described (See Carroll et al., J Steroid Biochem 21, 353-359 (1984) and Traish et al., J Steroid Biochem 23, 405-413 (1985)), their effectiveness is limited by the slow kinetics of dissociation of the endogenous receptor-bound androgen.

Another method used to assess receptor status is autoradiography. In this method disclosed in Barrack et al., "Current concepts and approaches to the study of prostate cancer," Progress in Clinical and Biological Research, 239, 155-187 (1987) the radioactively labeled androgen is incubated with slide-mounted tissue sections of small tissue biopsy specimens which are then frozen, sectioned and fixed. Nuclear localization of radioactivity is detected by exposure of tissue sections to x-ray film. This technique requires considerable technical expertise, is labor intensive, and requires extended periods of exposure time. It is therefore of limited usefulness in general clinical assays. Another problem is high levels of background signal, i.e. a high noise/signal ratio, making it difficult to distinguish receptor-bound nuclear radioactivity from unbound radioactivity distributed throughout the cells.

WO 87/05049 (Shine) discloses a method for the production of purified steroid receptor proteins, specifically estrogen receptor proteins, through the expression of recombinant DNA encoding for

such proteins in eukaryotic host cells. However, the reference does not disclose the sequence for androgen receptor protein, nor does it disclose a method for obtaining such a sequence.

5 SUMMARY OF THE INVENTION

The present invention provides a DNA sequence characterized by a structural gene coding for a polypeptide having substantially the same biological activity as androgen receptor protein. A DNA sequence encoding androgen receptor protein or a protein having
10 substantially the same biological activity as androgen receptor activity is also provided. DNA sequences may be obtained from cDNA or genomic DNA, or prepared using DNA synthesis techniques.

The invention further discloses cloning vehicles comprising a DNA sequence comprising a structural gene encoding a polypeptide
15 having substantially the same biological activity as androgen receptor protein. Cloning vehicles comprising a DNA sequence encoding androgen receptor protein or a protein having substantially the same biological activity as androgen receptor protein is also provided. The cloning vehicles further comprise a promoter sequence
20 upstream of and operatively linked to the DNA sequence. In general the cloning vehicles will also contain a selectable marker, and, depending on the host cell used, may contain such elements as regulatory sequences, polyadenylation signals, enhancers and RNA splice sites.

25 The invention further provides cells transfected or transformed to produce androgen receptor protein or a protein having substantially the same biological activity as androgen receptor protein.

30 A further aspect of the invention provides a purified androgen receptor protein and purified polypeptides and proteins having substantially the same biological activity as androgen receptor activity, and methods for producing such proteins and polypeptides.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a comparison of DNA-binding domains of the human androgen receptor (hAR) with members of the nuclear receptor family. (A) is a comparison of oligo A nucleotide sequence with sequences of hAR and other nuclear receptors: hPR, human progesterone receptor; hMR, human mineralocorticoid receptor; hGR, human glucocorticoid receptor; hER, human estrogen receptor; hT3R, human thyroid hormone receptor; hRAR, human retinoic acid receptor. Chromosomal locations are shown in parentheses at the left.

Nucleotide identity between oligo A and hAR is indicated with an asterisk. The percent homology with oligo A is in parentheses at the right of each sequence. (B) shows the structure of fibroblast clone ARHFL1 human fibroblast clone [1]). Nucleotide residues are numbered from the 5'-terminus. Restriction endonuclease sites were determined by mapping or were deduced from DNA sequence. The TGA translation termination codon, determined by comparison with hPR, hMR and hGR, follows a long open reading frame containing sequences homologous to those of other steroid receptors. Arrows indicate exon boundaries in genomic clone X05AR.

The hatched area is the putative DNA binding domain. (C) shows a comparison of amino acid sequences of the AR DNA-binding domain with sequences of the nuclear receptor family. AR amino acid sequence was deduced from nucleotide sequence of clone ARHFL1 and is numbered beginning with the first conserved cysteine residue (+). Amino acid numbers in parentheses at the left indicate the residue number of the first conserved cysteine from the references indicated above. Percent homology with hAR is indicated in parentheses on the right. The region of the DNA-binding domain from which the oligo A sequence was derived is underlined in hAR. Coding DNA of residues 1 to 31 is contained within genomic clone X05AR. Abbreviations in addition to those described above are cVDR, chicken vitamin D receptor, and VERBA, erb A protein from avian erythroblastosis virus.

Abbreviations for amino acid residues are:

A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

Figure 2 illustrates the steroid binding properties of expressed AR cDNA. (A) shows the structure of pCMVAR in the expression vector pCMV containing the human cytomegalovirus(CMV) promoter of the immediate early gene, poly(A) addition-transcription terminator region of the human growth hormone gene (hGH poly A), SV40 origin of replication (SV40 Ori), and a polylinker region for insertion of cDNAs. The plasmid pTEBR contains the ampicillin resistance gene (Amp). (B) shows saturation analysis of [³H]dihydrotestosterone binding in extracts of pCMVAR transection of COS M6 cells. Portions of cytosol (0.1 ml, 0.3 mg/ml protein) were incubated overnight at 4°C with increasing concentrations of ³H-labeled hormone and analyzed by charcoal adsorption. Nonspecific binding increased from 18% to 37% of total bound radioactivity. (C) shows a scratched plot analysis of [³H]dihydrotestosterone binding. Error estimation was based on linear regression analysis (r=0.966). (D) illustrates the competition of unlabeled steroids for binding of 5 nM [³H]dihydrotestosterone in transfected COS M6 cell extracts. Unlabeled steroids were added at 10- and 100-fold excess of labeled hormone. Specific binding was determined as previously described.

Figure 3 is a compiled clone map of the human androgen receptor. The map shows the structure of the human androgen receptor gene and the relative positions of the nucleic acid sequences contained in the cDNA probes [A], [B], [C] and [D], human fibroblast clone [1], human epididymis clones [1] and [5], human genomic clones [1], [2], [3], [4] and [5], and rat epididymis clones [1] and [2].

Figure 4 is a photograph showing chromosome localization of the AR gene on Southern blots of DNA from human cells containing multiple X chromosomes and mouse or hamster cells with X-autosome translocation chromosomes.

Figure 5 shows the complete single strand sequence (5085 bases) of the human androgen receptor and the deduced amino acid sequence. No intron sequence is included.

5 Figure 6 shows the complete single strand sequence (4260 bases) of the rat androgen receptor and the deduced amino acid sequence.

Figure 7 is a photograph of a frozen section of rat ventral prostate stained with antibodies (AR-52-3-p) to the AR peptide NH₂-Asp-His-Val-Leu-Pro-Ile-Asp-Tyr-Tyr-Phe-Pro-Pro-Gln-Lys-Thr in a dilution of 1 to 3000 using the avidin-biotin peroxidase
10 technique. Androgen receptor is indicated by brown staining of nuclei in epithelial cells.

Figure 8 is a photograph showing restriction fragment length polymorphisms in the human androgen receptor gene.

15 Figure 9 is a photograph showing a Southern blot analysis in the human androgen receptor gene in complete androgen insensitivity syndrome patients.

DETAILED DESCRIPTION OF THE INVENTION

In the description the following terms are employed:

20 Nucleotide

A monomeric unit of DNA or RNA consisting of a sugar moiety (pentose), a phosphate,, and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of the pentose) and that combination of base and sugar is a
25 nucleoside. The base characterizes the nucleotide. The four DNA bases are adenine ("A"), guanine ("G"), cytosine ("C") and thymine ("T"). The four RNA bases are A, G, C and uracil ("U"). A and G are purines, abbreviated to R, and C, T, and U are pyrimidines, abbreviated to Y.

30 DNA Sequence

A linear series of nucleotides connected one to the other by phosphodiester bonds between the 3' and 5' carbons of adjacent pentoses.

Codon

A DNA sequence of three nucleotides (a triplet) which encodes through mRNA an amino acid, a translational start signal or a translational termination signal. For example, the nucleotide triplets TTA, TTG, CTT, CTC, CTA and CTG encode for the amino acid leucine ("Leu"), TAG, TAA and TGA are translational stop signals and ATG is a translational start signal.

Reading Frame

The grouping of codon's during translation of mRNA into amino acid sequences. During translation the proper reading frame must be maintained. For example, the sequence GCTGGTTGTAAG may be translated in three reading frames or phases, each of which affords a different amino acid sequence:

GCT GGT TGT AAG - Ala-Gly-Cys-Lys
G CTG GTT GTA AG - Leu-Val-Val
GC TGG TTG TAA A - Trp-Leu-(STOP)

Polypeptide

A linear series of amino acids connected one to the other by peptide bonds between the α -amino and carboxy groups of adjacent amino acids.

Genome

The entire DNA of a substance. It includes inter alia the structural genes encoding for the polypeptides of the substance, as well as operator, promoter and ribosome binding and interaction sequences including sequences such as the Shine-Dalgarno sequences.

Structural Gene

A DNA sequence which encodes through its template or messenger RNA ("mRNA") a sequence of amino acids characteristic of a specific polypeptide.

Transcription

The process of producing mRNA from a structural gene.

Translation

The process of producing a polypeptide from mRNA.

Expression

The process undergone by a structural gene to produce a polypeptide. It is a combination of transcription and translation.

Plasmid

5 A non-chromosomal double-stranded DNA sequence comprising an intact "replicon" such that the plasmid is replicated in a host cell. When the plasmid is placed within a unicellular organism, the characteristics of that organism are changed or transformed as a result of the DNA of the plasmid. For example, a plasmid carrying
10 the gene for tetracycline resistance (Tet^R) transforms a cell previously sensitive to tetracycline into one which is resistant to it. A cell transformed by a plasmid is called a "transformant".

Phage or Bacteriophage

Bacterial virus many of which include DNA sequences
15 encapsidated in a protein envelope or coat ("capsid"). In a unicellular organism a phage may be introduced as free DNA by a process called transfection.

Cloning Vehicle

A plasmid, phage DNA or other DNA sequences which are able to
20 replicate in a host cell, characterized by one or a small number of endonuclease recognition sites at which such DNA sequences may be cut in a determinable fashion without attendant loss of an essential biological function of the DNA, e.g., replication, production of coat proteins or loss of promoter or binding sites, and which
25 contain a marker suitable for use in the identification of transformed cells, e.g., tetracycline resistance or ampicillin resistance. A cloning vehicle is often called a vector.

Cloning

The selection and propagation of a single species.

30 Recombinant DNA Molecule

A hybrid DNA sequence comprising at least two nucleotide sequences, the first sequence not normally being found together in nature with the second.

Expression Control Sequence

A DNA sequence of nucleotides that controls and regulates expression of structural genes when operatively linked to those genes.

5 To attain the objects of this invention it was necessary to determine the amino acid sequence and the DNA sequence of the structural gene encoding androgen receptor protein. One conventional approach would involve starting with the purified androgen receptor protein. However, as described above, significant
10 amounts of the protein for such purposes have not been obtained.

An alternative approach to circumvent the overwhelming difficulty of androgen receptor protein purification is direct isolation of the DNA encoding the messenger RNA for androgen receptor protein.

15 Our strategy for isolating AR DNA was based on evidence that the AR gene is X-linked and that no other steroid receptor gene is located on the X chromosome. Sequence data are available from cDNAs for glucocorticoid, estrogen, progesterone, mineralocorticoid and vitamin D receptors. Comparison of the derived amino acid sequences
20 has revealed a central region of high cysteine content which was found also in the v-erb A oncogene product recently identified as the thyroid hormone receptor. Within this 61-63 amino acid region is an arrangement of 9 cysteine residues that are absolutely conserved among steroid receptors thus far characterized. The
25 overall homology among sequences in this conserved region ranges between 40 and 90%. We assumed that AR would resemble other members of the steroid receptor family in the conserved DNA-binding domain.

A human X chromosomal library was screened with the synthetic oligo nucleotide probe A (Oligo A sequence = ^{5'} CTT TTG AAG AAG
30 ACC TTA CAG CCC TCA CAG GT^{3'}) of Figure 1 (A) designed as a consensus sequence from the conserved sequence of the DNA-binding domain of other steroid receptors. Screening the library with the oligo A probe resulted in several recombinants whose inserts were cloned into bacteriophage M13 DNA and sequenced. One recombinant

clone (Charon 35 X05AR) (human genomic clone [1]) contained a sequence similar to, yet distinct from, the DNA-binding domains of other steroid receptors. It had 84% sequence identity with oligo A, while other receptor DNAs were 78% to 91% homologous with the consensus oligonucleotide.

From the nucleotide sequence just 5' of the DNA binding domain, oligonucleotide probe B (Oligo B sequence = ⁵GGA CCA TGT TTT GCC CAT TGA CTA TTA CTT TCC ACC CC^{3'}) was synthesized and used to screen bacteriophage lambda gt11 cDNA libraries from human epididymis and cultured human foreskin fibroblasts. Recombinant phage (unamplified) screened with this oligonucleotide by in situ hybridization revealed one positive clone in each library. The epididymal clone (gt11 ARHELL)(human epididymis clone [1]) contained the complete DNA-binding domain and approximately 1.5 kb of upstream sequence, whereas the fibroblast clone (gt11 ARHFL1)(human fibroblast clone [1]) shown in Figure 1(B) contained the DNA-binding domain and 1.5 kb of downstream sequence. The DNA-binding domains of the cDNA isolates were identical to that of the genomic exon sequence.

Transient expression in monkey kidney cells (COS M6) demonstrated that the human foreskin fibroblast cDNA fragment encodes the steroid-binding domain of hAR. A DNA fragment (ARHFLIH-X) extending 5' to 3' from the Hind III site within the putative DNA-binding domain through the stop codon (TGA) was cloned into pCMV as shown in Figure 2(A). Expression was facilitated by adding to the 5' end a consensus translation initiation sequence containing the methionine codon (ATG) in reading frame.

Transfection of the recombinant construct produced a protein with high-affinity for [³H]dihydrotestosterone, Figure 2(C) saturable at physiological levels of hormone. See Figure 2(B). The binding constant [$K_d = 2.7 (+ 1.4) \times 10^{-10} M$] was nearly identical to that of native AR. The level of expressed protein, 1.3 pmol per milligram of protein, was 20 to 60 times greater than that in male reproductive tissues. Mock transfections without plasmid or

transfections with plasmid DNA lacking the AR insert yielded no specific binding of dihydrotestosterone. Figure 2(D) shows steroid specificity was identical to that of native AR, with highest affinity for dihydrotestosterone and testosterone, intermediate
5 affinity for progesterone and estradiol, and low affinity for cortisol.

Figure 3 is a clone map compiled to show the human androgen receptor gene and the nucleic acid sequences in the cDNA clones, human genomic clones, human fibroblast clones, human epididymis
10 clones, and rat epididymis clones. Human fibroblast clone [1] extended through the stop codon or the C-terminal end of the androgen receptor protein. To isolate and elucidate the sequence of the 5' or N-terminal end of the androgen receptor protein, we used a EcoRI/SstI fragment (EcoRI site was from the linker) from the 5' end
15 of human epididymis clone [1] as a probe (cDNA probe [A]), to rescreen the human X chromosomal library by standard techniques. By these techniques, human genomic clone [2] was isolated and in turn used as a probe to rescreen a human epididymis library and isolate human epididymis clone [5]. The N-terminal sequence was elucidated
20 along with the 5' flanking sequence of the androgen receptor protein and gene. Human genomic clones [3], [4] and [5] for the sequence 3' of human genomic clone [1] were obtained using cDNA probes B [a Hind III/EcoRI fragment] and C [an EcoRI fragment], by screening and isolating by standard techniques.

Two rat clones, rat epididymis clones [1] and [2], were
25 isolated from a rat epididymis cDNA library using as probes the complete human epididymis clone [1] and a EcoRI/PstI fragment, cDNA probe [D], respectively. These rat clones contained the entire protein coding sequence for the rat androgen receptor, plus flanking
30 5' and 3' untranslated sequences which were used to confirm the sequence of the human androgen receptor.

The complete double-stranded sequence encoding the human androgen receptor protein was determined and is set forth in Figure 4. The single-stranded DNA sequence encoding human androgen

receptor protein along with the amino acid sequence which it codes for are set forth in Figure 5. The single stranded DNA sequence and the amino acid sequence for the rat androgen receptor protein is set forth in Figure 6

5 Recombinant DNA clones human fibroblast clone [1] isolated from human foreskin fibroblast cDNA gt11 expression library, human epididymis clones [1] and [5] isolated from human epididymis cDNA gt11 expression library were deposited in the American Type Culture Collection with accession numbers ATCC # _____, ATCC # _____ and
10 ATCC # _____ respectively. Human genomic clones [1], [2], [3], [4] and [5] which were isolated from human X chromosome lambda Charon 35 library available as ATCC # 57750 have been deposited with the American Type Culture Collection with accession numbers ATCC
15 # _____, ATCC # _____, ATCC # _____, ATCC # _____ and ATCC # _____ respectively.

 A wide variety of host=cloning vehicle combinations may be usefully employed in cloning the double stranded DNA disclosed herein. For example, useful cloning vehicles may include
20 chromosomal, non-chromosomal and synthetic DNA sequences such as various known bacterial plasmids and wider host range plasmids such as pCMV and vectors derived from combinations of plasmids and phage DNA such as plasmids which have been modified to employ phage DNA expression control sequences. Useful hosts may include bacterial
25 hosts, yeasts and other fungi, animal or plant hosts, such as Chinese Hamster Ovary cells (CHO, or monkey kidney cells (COS M6), and other hosts. The particular selection of host-cloning vehicle combinations may be made by those of skill in the art after due consideration of factors such as the source of the DNA- i.e. genomic or cDNA.

30 Cloning vehicles for use in carrying out the present invention will further comprise a promoter operably linked to the DNA sequence encoding the androgen receptor protein. In some instances it is preferred that cloning vehicles further comprise an origin of

replication , as well as sequences which regulate and/or enhance expression levels, depending on the host cell selected.

Techniques for transforming hosts and expressing foreign cloned in them are well known in the art (See, for example, Maniatis et al., infra). Cloning vehicles used for expressing foreign genes in bacterial hosts will generally contain a selectable marker, such as a gene for antibiotic resistance, and a promoter which functions in the host cell.

Eukaryotic microorganisms, such as the yeast *Saccharomyces cerevisiae*, may also be used as host cells. Cloning vehicles will generally comprise a selectable marker, such as the nutritional marker TRP, which allows selection in a host strain carrying a *trp1* mutation. To facilitate purification of an androgen receptor protein produced in a yeast transformant, a yeast gene encoding a secreted protein may be joined to the sequence encoding androgen receptor protein.

Higher eukaryotic cells can also serve as host cells in carrying out the present invention. Cultured mammalian cells are preferred. Cloning vehicles for use in mammalian cells will comprise a promoter capable of directing the transcription of a foreign gene introduced into a mammalian cell. Also contained in the expression vector is a polyadenylation signal, located downstream of the insertion site. The polyadenylation signal can be that of the cloned androgen receptor gene, or may be derived from a heterologous gene.

A selectable marker, such as a gene that confers a selectable phenotype, is generally introduced into the cells along with the gene of interest. Preferred selectable markers include genes that confer resistance to drugs, such as neomycin, hygromycin and methotrexate. Selectable markers may be introduced into the cell on a separate plasmid at the same time as the gene of interest, or they may be introduced on the same plasmid.

The copy marker of the integrated gene sequence can be increased through amplification by using certain selectable

markers. Through selection, expression levels may be substantially increased.

• Androgen receptor proteins may be purified from the host cells or cell media according to the present invention using techniques well known to those in the art. Such proteins may be utilized to produce monoclonal or polyclonal antibodies according to the techniques described below.

The techniques of this invention offer considerable advances over existing technology for measurement of androgen receptor. Utilizing proteins and peptides containing the disclosed sequences monoclonal or polyclonal antibodies can be produced for use as immunochemical reagents in immunodiagnostic assays. For example, radioimmunoassays and ELISA assays can be developed utilizing these reagents which will allow detection and quantification of androgen receptor in the presence of endogenous androgen since such androgen will not interfere with antibody binding to the receptor.

Immunocytochemistry utilizing our reagents enables determination and quantification of the cellular distribution of the androgen receptor in tumor tissues, which are often heterogenous in composition. This assay offers great potential for diagnostic evaluation of prostate cancer to determine responsiveness to androgen withdrawal therapy.

In addition, the antibodies produced using the disclosed amino acid sequences can also be used in processes for the purification of androgen receptor protein produced by the above methods. One such purification process is disclosed in Logeat, F., et al., Biochemistry vol. 24 (1985), pp. 1029-1035, which is incorporated by reference herein.

Androgen receptor proteins and polypeptides synthesized from the deduced amino acid sequence can be used as immunogens for the preparation of antibodies to the androgen receptor. Peptides for such use range in length from about 3 to about 958 amino acids in length and are preferably from about 15 to about 30 amino acids in length. Shorter peptides may have significant sequence homology to

other steroid receptor proteins and larger peptides may contain multiple antigenic determinants; these properties could result in antibodies with cross-reactivities to other steroid receptor proteins.

5 Peptides can be synthesized from amino acid sequences in the NH_2 -terminal region, the DNA-binding domain, and the carboxyl-terminal steroid binding domain. Peptide selection will be based on hydropathic plots, selecting hydrophilic regions that are more likely exposed on the receptor surface. For diagnostic
10 purposes preferred sequences will be selected from the NH_2 -terminal region where there is the least homology with other steroid receptor proteins.

15 Peptides for use as immunogens can be synthesized using techniques available to one of ordinary skill in the art. For example, peptides corresponding to androgen receptor sequences can be synthesized using tBOC chemistry on a Biosearch Model 9500 peptide synthesizer. Peptide purity is assessed by high pressure liquid chromatography. Peptides can be conjugated to keyhole limpet hemocyanin through cysteine residues using the coupling agent
20 m-maleimido-benzoyl-N-hydroxysuccinimide ester. One can also prepare resin-bound peptides utilizing the p-(oxymethyl benzamide) handle to attach the C-terminal amino acid to solid-phase resin support.

25 Proteins and peptides of this invention can be utilized for the production of polyclonal or monoclonal antibodies. Methods for production of such antibodies are known to those of ordinary skill in the art and may be performed without undue experimentation. One method for the production of monoclonal antibodies is described in Kohler, G., et al., "Continuous Culture of Fused Cells Secreting
30 Antibody of Predefined Specificity," Nature, vol. 256 (1975), p. 495, which is incorporated herein by reference. Polyclonal antibodies, by way of example, can be produced by the method described below.

Peptide conjugates or resin-bound peptides can be injected into rabbits according to the procedure of Vaitukaitis et al., J Clin Endocrinol Metab, 33, 988-991 (1971) using a standard immunization schedule. Antisera titers can be determined in the ELISA assay.

5 For example, one androgen receptor sequence,
NH₂-Asp-His-Val-Leu-Pro-Ile-Asp-Tyr-Tyr-Phe-Pro-Pro-Gln-Lys-Thr
in the 5' region upstream from the DNA-binding domain, was used to
raise antisera in rabbits. The antisera react selectively at a
dilution of 1 to 500 with the androgen receptor both in its
10 untransformed 8-10S form and in its 4-5S transformed form. Receptor
sedimentation on sucrose gradients increases from 4 to 8-10S in the
presence of antiserum at high ionic strength and from 8-10S to
11-12S at low ionic strength sucrose gradients. In the ELISA
reaction against the peptide used as immunogen, reactivity was
15 detectable at 1 to 25,000 dilution. This antiserum at a dilution of
1 to 3000 was found effective in staining nuclear androgen receptor
in rat prostate and other male accessory sex glands (see Figure 7).

Our invention provides new molecular probes comprising
complementary DNA sequences derived from the deduced sequences
20 encoding the androgen receptor for diagnostic purposes. Such probes
may be used to detect the presence of androgen receptor mRNA in
tumor cells. Such probes may also be used for detection of androgen
receptor gene defects. Androgen receptor complementary DNA
sequences can be used as hybridization probes to detect
25 abnormalities in the androgen receptor gene or in its messenger RNA.

Androgen receptor DNA sequences disclosed and complementary RNA
sequences can be used to construct probes for use in DNA
hybridization assays. An example of one such hybridization assay
and methods for constructing probes for such assays are disclosed
30 in U.S. Patent No. 4,683,195 to Mullis et al., U.S. Patent No.
4,683,202 to Mullis, U.S. Patent No. 4,617,261 to Sheldon, III et
al., U.S. Patent No. 4,683,194 to Salki et al., and U.S. Patent No.
4,705,886 to Levenson et al., which are hereby incorporated by
reference.

By example, one method for detecting gene deletion utilizes Southern blotting and hybridization. DNA can be isolated from cultured skin fibroblasts or from leukocytes obtained from blood. DNA is cut with restriction enzymes, electrophoresed on an agarose gel, blotted onto nitrocellulose, and hybridized with [32P]-labeled androgen receptor DNA (see Maniatis, T. et al., Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, 1982, incorporated by reference herein).

In addition, small mutations can be detected utilizing methods known to one of ordinary skill in the art, from cultured skin fibroblasts of the affected individual. A cDNA library can be prepared using standard techniques. The androgen receptor clones can be isolated using a [32P]DNA AR probe. The clones AR cDNA can then be sequenced and compared to normal AR cDNA sequences.

Alternatively genomic DNA can be isolated from blood leukocytes or cultured skin fibroblasts of the affected individual. The DNA is then subjected to restriction enzyme digestion, electrophoresis and is blotted onto nitrocellulose. Synthetic oligonucleotides can be used to bracket specific exons. Exon sequences are amplified using the polymerase chain reaction, cloned into M13 and sequenced. The sequences are compared to normal human AR DNA sequences.

Another method of identifying small mutations or deletions takes advantage of the ability of RNase A to cleave regions of single stranded RNA in RNA:DNA hybrids. Genomic DNA isolated from fibroblasts of affected individuals is hybridized with radioactive RNA probes (Promega Biotec) prepared from wild-type androgen receptor cDNA. Mismatches due to mutations would be cleaved by RNase A and result in altered sized bands relative to wild-type on denaturing polyacrylamide gels.

Restriction fragment length polymorphism (RFLP) linked to the androgen receptor gene locus may be used in prenatal diagnosis and carrier detection of androgen insensitivity. For example, the presence of RFLPs in normal individuals is first established by isolating DNA from lymphocytes of at least six females (total of 12

X chromosomes). DNA can be isolated using the proteinase K procedure and fragmented using a battery of restriction enzymes. Preferred are those enzymes that contain the dinucleotide sequence CG in their recognition sequence. Southern blots are screened with 5-10 kb androgen receptor genomic fragments which if possible lack repetitive DNA. For those regions containing repetitive elements, total human genomic DNA can be added as competitor in the hybridization reaction. Alternatively, one can subclone selected regions to yield a probe free of repetitive elements.

For example, a human restriction fragment length was determined by cDNA probe (B) and Hind III restriction endonuclease using the Southern blot technique (See Figure 8). The two RFLP alleles detected are a fragment at 6.5 kb (allele) and a fragment at 3.5 kb (allele 2). Major constant fragment bands are seen at approximately 2 and 5 kb with minor constant bands at 0.9 and 7.5 kb. Allele 1 is present in approximately 30% of the X chromosomes of the Caucasian population. Allele 2 is present in approximately 20% of the X chromosomes of the Caucasian population. In Figure 9 Lanes A, B and D, DNA from women who are homozygous for allele 1 is shown. In Figure 9 Lane C, DNA from a woman who is heterozygous for both alleles 1 and 2 is shown. Figure 9 Lane E contains DNA from a man that only possesses allele 2. This RFLP, and others determined by the clones we have isolated, will enable one to monitor the androgen receptor gene in various disease conditions described herein.

An example of using the androgen receptor clones to detect mutations is shown in Figure 9 where five different complete androgen insensitive patients' DNA are digested with EcoRI, electrophoresed on a Southern blot, and probed with cDNA probe B. The patient in lane B lacks a 3kb band indicating that part of the androgen receptor gene is deleted. Further analysis of this and other patients DNA is possible with other AR probes and by sequencing by standard methods and comparing the abnormal sequence to the normal sequence described herein.

Other potential uses for oligonucleotide sequences disclosed, for example in construction of therapeutics to block genetic expression, will be obvious to one of ordinary skill in the art.

5

10

15

20

25

30

What is claimed is:

1. A recombinant DNA molecule comprising a DNA sequence encoding the structural gene for androgen receptor protein.

2. The recombinant DNA molecule of Claim 1 wherein the androgen receptor protein is a human androgen receptor protein.

3. A cloning vehicle comprising a genomic DNA molecule which upon expression in a eukaryotic host produces androgen receptor protein.

4. The cloning vehicle of Claim 3 wherein the androgen receptor protein is a human androgen receptor protein.

5. An androgen receptor protein produced by translation of the DNA sequence encoding androgen receptor protein in a host organism transfected or transformed by the cloning vehicle of Claim 3.

6. A human androgen receptor protein produced by translation of the DNA sequence encoding human androgen receptor protein in a host organism transfected or transformed by the cloning vehicle of Claim 4.

ABSTRACT OF THE DISCLOSURE

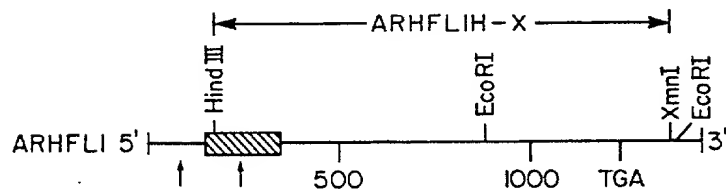
DNA sequences encoding human androgen receptor protein and polypeptides and proteins having substantially the same biological activity as human androgen receptor protein and the amino acid sequences of human androgen receptor protein and polypeptides and proteins having substantially the same biological activity as human androgen receptor protein are disclosed. Methods for the production and use of such compositions are also disclosed.

FIGURE 1

A.

Oligo A	Complement	5'-ACC	TGT	GAG	GGC	TGT	AAG	GTC	TTC	TTC	AAA	AG-3'	(100%)
		**	***	*	*	**	***	***	***	***	***	**	
hAR (X)		ACA	TGT	GGA	AGC	TGC	AAG	GTC	TTC	TTC	AAA	AG	(84%)
hPR (11)		ACC	TGT	GGG	AGC	TGT	AAG	GTC	TTC	TTT	AAG	AG	(88%)
hMR (4)		ACC	TGT	GGC	AGC	TGC	AAA	GTT	TTC	TTC	AAA	AG	(81%)
hGR (5)		ACT	TGT	GGA	AGC	TGT	AAA	GTT	TTC	TTC	AAA	AG	(81%)
hER (6)		TCC	TGT	GAG	GGC	TGT	AAG	GCC	TTC	TTC	AAG	AG	(91%)
hT3R (3, 17)		ACG	TGT	GAA	GGC	TGC	AAG	GGT	TTC	TTT	AGA	AG	(78%)
hRAR (17)		GCC	TGT	GAG	GGC	TGC	AAG	GGC	TTC	TTC	CGC	CG	(78%)

B.



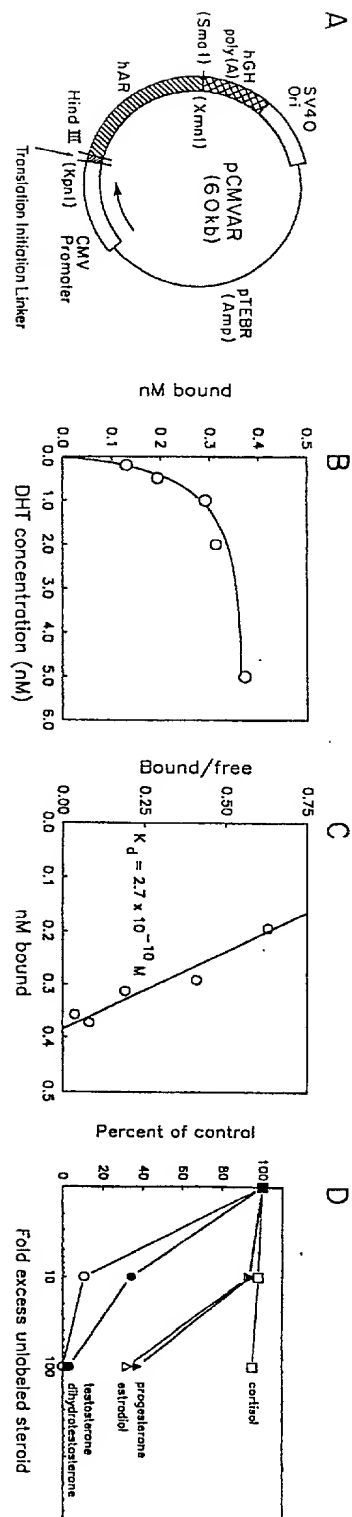
C.

DNA-Binding Domain

		+	+	10										+	20+										30										
hAR		C	L	I	C	G	D	E	A	S	G	C	H	Y	G	A	L	<u>T</u>	C	G	S	<u>C</u>	<u>K</u>	<u>V</u>	<u>F</u>	<u>F</u>	<u>K</u>	<u>R</u>	A	A	E	G	(100%)		
hPR	(aa 567)	C	L	I	C	G	D	E	A	S	G	C	H	Y	G	V	L	T	C	G	S	C	K	V	F	F	F	K	R	A	M	E	G	(94%)	
hMR	(aa 603)	C	L	V	C	G	D	E	A	S	G	C	H	Y	G	V	V	T	C	G	S	C	K	V	F	F	F	K	R	A	V	E	G	(87%)	
hGR	(aa 421)	C	L	V	C	S	D	E	A	S	G	C	H	Y	G	V	L	T	C	G	S	C	K	V	F	F	F	K	R	A	V	E	G	(87%)	
hER	(aa 185)	C	A	V	C	N	D	Y	A	S	G	Y	H	Y	G	V	W	S	C	E	G	C	K	A	F	F	F	K	R	S	I	Q	G	(55%)	
cVDR		C	G	V	C	G	D	R	A	T	G	F	H	F	N	A	M	T	C	E	G	C	K	G	F	F	F	R	R	S	M	K	R	(48%)	
hT3R	(aa 102)	C	V	V	C	G	D	K	A	T	G	Y	H	Y	R	C	I	T	C	E	G	C	K	G	F	F	F	R	R	T	I	Q	K	(48%)	
vERBA	(aa 37)	C	V	V	C	G	D	K	A	T	G	Y	H	Y	R	C	I	T	C	E	G	C	K	S	F	F	F	R	R	T	I	Q	K	(48%)	
hRAR	(aa 58)	C	F	V	C	Q	D	K	S	S	G	Y	H	Y	G	V	S	A	C	E	G	C	K	G	F	F	F	R	R	S	I	Q	K	(45%)	

			+	40										+	50										+	+	60+										
hAR		K	Q	K	Y	L	C	A	S	R	N	D	C	T	I	D	K	F	R	R	K	N	C	P	S	C	R	L	R	K	C	Y	E	A	G	M	(100%)
hPR		Q	<u>H</u>	N	Y	L	C	A	G	R	N	D	C	I	V	D	K	I	R	R	K	N	C	P	A	C	R	L	R	K	C	C	Q	A	G	M	(71%)
hMR		Q	<u>H</u>	N	Y	L	C	A	G	R	N	D	C	I	I	D	K	I	R	R	K	N	C	P	A	C	R	L	Q	K	C	L	Q	A	G	M	(71%)
hGR		Q	<u>H</u>	N	Y	L	C	A	G	R	N	D	C	I	I	D	K	I	R	R	K	N	C	P	A	C	R	Y	R	K	C	L	Q	A	G	M	(71%)
hER		<u>H</u>	N	D	Y	M	C	P	A	T	N	Q	C	T	I	D	K	N	R	R	K	S	C	Q	A	C	R	L	R	K	C	Y	E	V	G	M	(63%)
cVDR		K	A	M	F	T	C	P	F	N	G	D	C	K	I	T	K	D	N	R	R	H	C	Q	A	C	R	L	K	R	C	V	D	I	G	M	(40%)
hT3R	N L	<u>H</u>	P	S	Y	S	C	K	Y	E	G	K	C	V	I	D	K	V	T	R	N	Q	C	Q	E	C	R	F	K	K	C	I	Y	V	G	M	(40%)
vERBA	N L	<u>H</u>	P	T	T	S	C	T	Y	D	G	C	C	V	I	D	K	I	T	R	N	Q	C	Q	L	C	R	F	K	K	C	I	S	V	G	M	(37%)
hRAR		N	M	V	Y	T	C	H	R	D	K	N	C	I	I	N	K	V	T	R	N	R	C	Q	Y	C	R	L	Q	K	C	F	E	V	G	M	(43%)

FIGURE 2



Five



2000

Variable	Mean	SD	Min	Max
Age	34.5	10.2	18	65
Gender	Male	10.5	0	20
Gender	Female	19.5	0	20
Marital status	Married	15.5	0	20
Marital status	Single	4.5	0	10
Marital status	Divorced	0.5	0	1
Marital status	Widowed	0.5	0	1
Education	High school	10.5	0	20
Education	College	10.5	0	20
Education	Postgraduate	0.5	0	1
Occupation	Manager	10.5	0	20
Occupation	Professional	10.5	0	20
Occupation	Service	10.5	0	20
Occupation	Unemployed	0.5	0	1
Income	Low	10.5	0	20
Income	Medium	10.5	0	20
Income	High	0.5	0	1
Health status	Good	10.5	0	20
Health status	Fair	10.5	0	20
Health status	Poor	0.5	0	1
Stress level	Low	10.5	0	20
Stress level	Medium	10.5	0	20
Stress level	High	0.5	0	1
Life satisfaction	High	10.5	0	20
Life satisfaction	Medium	10.5	0	20
Life satisfaction	Low	0.5	0	1

Variable	Mean	SD	Min	Max
Age	38.5	12.5	18	65
Gender	0.5	0.5	0	1
Marital status	0.5	0.5	0	1
Education	12.5	2.5	8	16
Income	15.5	5.5	10	25
Health status	0.5	0.5	0	1
Smoking status	0.5	0.5	0	1
Alcohol consumption	0.5	0.5	0	1
Exercise frequency	0.5	0.5	0	1
Stress level	0.5	0.5	0	1
Sleep quality	0.5	0.5	0	1
Work satisfaction	0.5	0.5	0	1
Life satisfaction	0.5	0.5	0	1
Depression score	0.5	0.5	0	1
Anxiety score	0.5	0.5	0	1
Quality of life score	0.5	0.5	0	1
Healthcare utilization	0.5	0.5	0	1
Health insurance status	0.5	0.5	0	1
Chronic disease status	0.5	0.5	0	1
Medication adherence	0.5	0.5	0	1
Healthcare provider satisfaction	0.5	0.5	0	1
Healthcare system trust	0.5	0.5	0	1
Healthcare access	0.5	0.5	0	1
Healthcare cost	0.5	0.5	0	1
Healthcare quality	0.5	0.5	0	1
Healthcare equity	0.5	0.5	0	1
Healthcare transparency	0.5	0.5	0	1
Healthcare accountability	0.5	0.5	0	1
Healthcare innovation	0.5	0.5	0	1
Healthcare sustainability	0.5	0.5	0	1
Healthcare resilience	0.5	0.5	0	1
Healthcare leadership	0.5	0.5	0	1
Healthcare governance	0.5	0.5	0	1
Healthcare regulation	0.5	0.5	0	1
Healthcare monitoring	0.5	0.5	0	1
Healthcare evaluation	0.5	0.5	0	1
Healthcare improvement	0.5	0.5	0	1
Healthcare reform	0.5	0.5	0	1
Healthcare development	0.5	0.5	0	1
Healthcare progress	0.5	0.5	0	1
Healthcare achievement	0.5	0.5	0	1
Healthcare success	0.5	0.5	0	1
Healthcare impact	0.5	0.5	0	1
Healthcare contribution	0.5	0.5	0	1
Healthcare legacy	0.5	0.5	0	1
Healthcare vision	0.5	0.5	0	1
Healthcare mission	0.5	0.5	0	1
Healthcare values	0.5	0.5	0	1
Healthcare principles	0.5	0.5	0	1
Healthcare beliefs	0.5	0.5	0	1
Healthcare attitudes	0.5	0.5	0	1
Healthcare behaviors	0.5	0.5	0	1
Healthcare outcomes	0.5	0.5	0	1
Healthcare results	0.5	0.5	0	1
Healthcare performance	0.5	0.5	0	1
Healthcare excellence	0.5	0.5	0	1
Healthcare innovation	0.5	0.5	0	1
Healthcare sustainability	0.5	0.5	0	1
Healthcare resilience	0.5	0.5	0	1
Healthcare leadership	0.5	0.5	0	1
Healthcare governance	0.5	0.5	0	1
Healthcare regulation	0.5	0.5	0	1
Healthcare monitoring	0.5	0.5	0	1
Healthcare evaluation	0.5	0.5	0	1
Healthcare improvement	0.5	0.5	0	1
Healthcare reform	0.5	0.5	0	1
Healthcare development	0.5	0.5	0	1
Healthcare progress	0.5	0.5	0	1
Healthcare achievement	0.5	0.5	0	1
Healthcare success	0.5	0.5	0	1
Healthcare impact	0.5	0.5	0	1
Healthcare contribution	0.5	0.5	0	1
Healthcare legacy	0.5	0.5	0	1
Healthcare vision	0.5	0.5	0	1
Healthcare mission	0.5	0.5	0	1
Healthcare values	0.5	0.5	0	1
Healthcare principles	0.5	0.5	0	1
Healthcare beliefs	0.5	0.5	0	1
Healthcare attitudes	0.5	0.5	0	1
Healthcare behaviors	0.5	0.5	0	1
Healthcare outcomes	0.5	0.5	0	1
Healthcare results	0.5	0.5	0	1
Healthcare performance	0.5	0.5	0	1
Healthcare excellence	0.5	0.5	0	1

10 GAGCTCTGGA CTCGAGACCT	20 CAAAATTGAG GTTTTAACTC	30 GCGCTATGTG GCGGATACAC	40 TACATGGCAA ATGTACCGTT	50 GTGTTTTTAG CACAAAAATC	60 TGTTTGTGTG ACAAACACAC
70 TTTACCTGCT AAATGGACGA	80 TGTCTGGGTG ACAGACCCAC	90 ATTTTGCCTT TAAAACGGAA	100 TGAGAGTCTG ACTCTCAGAC	110 GATGAGAAAT CTACTCTTTA	120 GCATGGTTAA CGTACCAATT
130 AGGCAATTCC TCCGTTAAGG	140 AGACAGGAAG TCTGTCTTTC	150 AAAGGCAGAG TTTCCGTCTC	160 AAGAGGGTAG TTCTCCCATC	170 AAATGACCTC TTTACTGGAG	180 TGATTCTTGG ACTAAGAACC
190 GGCTGAGGGT CCGACTCCCA	200 TCCTAGAGCA AGGATCTCGT	210 AATGGCACAA TTACCGTGTT	220 TGCCACGAGG ACGGTGCTCC	230 CCCATCTAT GGGCTAGATA	240 CCCTATGACG GGGATACTGC
250 GAACTCTAAG CTTGAGATTG	260 GTTTCAGCAT CAAAGTCGTA	270 CAGCTATCTG GTCGATAGAC	280 CTGGCTTGGT GACCGAACCA	290 CACTGGCTTG GTGACCGAAC	300 CCTCCTCAGT GGAGGAGTCA
310 TTGTAGGAGA AACATCCTCT	320 CTCTCCCACT GAGAGGGTGA	330 CTCCCATCTG GAGGGTAGAC	340 GCGGCTCTTA GCGCBAGAAAT	350 TCAGTCCTGA AGTCAGGACT	360 AAAGAACCCN TTTCTTGGGN
370 TGGCNAGCCA ACCGNTCGGT	380 GGAGCNAGGT CCTCGNTCCA	390 ATTGNTATCG TAAGNATAGC	400 TCCTTTTCNT AGGAAAAGNA	410 CCTCCTNGCC GGAGGANCGB	420 TCACCTNGTT AGTGGANCAA
430 GNTTTTTTAG CNAAAAAATCT	440 TTGGNCTTNG AACCNGAANC	450 NAACCAAATT NTTGGTTTAA	460 TGTATGCTGG ACATACGACC	470 CCTCCAGGAA GGAGGTCCTT	480 ATCTGGAGCC TAGACCTCGG
490 TGCGCGCTAA ACCGGCGATT	500 ACCTTGGTTT TGGAAACCAA	510 AGGAAAGCAG TCCTTTCGTC	520 GAGCTATTCA CTCGATAAGT	530 GGAAGCAGGG CCTTCGTCCC	540 TCCTCCAGGG AGGAGGTCCT
550 CTAGAGCTAG GATCTCGATC	560 CCTCTCCTGC GGAGAGGACG	570 CCTCGCCAC GGAGCGGGTG	580 GTGCGCCAGC CACGCGGTG	590 ACTTGTTTCT TGAACAAAGA	600 CCAAAGCNAC GGTTTCGNTG
610 TAGGCAGGCG ATCCGTCCGC	620 TTAGCGCGCG AATCGCGCGC	630 GTGAGGGGAG CACTCCCTC	640 GGGAGAAAAG CCCTCTTTTC	650 GAAAGGGGAG CTTTCCCTC	660 GGGAGGGGAA CCCTCCCTTT
670 AGGAGGTGGG TCCTCCACCC	680 AAGGCAAGGA TTCCGTTCCCT	690 GGCCGGCCNG CCGGCCGGNC	700 GTGGGGGGCG CAGCCCCGCC	710 GACCCGACTC CTGGGCTGAG	720 GCANNAACTG CGTNNTTGAC
730 TTGCATTTGC AACGTAAACG	740 TCTCCACCTC AGAGGTGGAG	750 CCAGCGCCCC GGTCGCGGGG	760 CTCCGAGATC GAGGCTCTAG	770 CCGGGGAGCC GGCCCCCTCG	780 AGCTTGCTGG TCGACGACC
790 GAGAGCGGGA CTCTCGCCCT	800 ACGGTCCGGA TGCCAGGCCCT	810 GCAAGCCAG CGTTCGGGTC	820 AGGCAGAGGA TCCGTCTCCT	830 GGCBACAGAG CCGCTGTCTC	840 GGAAAAAGGG CCTTTTTCCC
850 CCCNAGCTAG GGGNTCGATC	860 CCGCTCCAGT GGCGAGGTCA	870 GCTGTACAGN CGACATGTGN	880 AGCCGAAGGA TCGGCTTCCT	890 CGCACCACGC GCCTGGTGGC	900 CAGCCCCAGC GTCGGGGTCC

FIGURE 4 (page 2 of 6)

910	920	930	940	950	960
CCGGCTCCAG	CGACAGCNAA	CGCCTCTTGC	ANGCSTTCGA	AGCCGCGGCC	CGGAGCTGCC
GGCCGAGGTC	GCTGTCGNTT	GGGAGAAACG	TNCGCAAGCT	TCGGCGGCGG	GCCTCGACGG
970	980	990	1000	1010	1020
CTTTCCTCTT	CGGTGAAGTT	TTTAAAGACT	GCTAAAGACT	CGGAGGAAGC	AAGGAAAGTG
GAAAGGAGAA	GCCACTTCAA	AAATTTTCGA	CGATTTCTGA	GCCTCCTTCG	TTCTTTTCAC
1030	1040	1050	1060	1070	1080
CCTGGTAGGA	CTGACGGCTG	CCTTTGTCCT	CCTCCTCTCC	ACCCGCGCTC	CCCCCACCCT
GGACCATCCT	GACTGCCGAC	GGAAACAGGA	GGAGGAGAGG	TGGGGCGGAG	GGGGGTGGGA
1090	1100	1110	1120	1130	1140
GCCTTCCCCC	CCTCCCCCGT	CTTCTCTCCC	GCAGCTGCCCT	CAGTCGGCTA	CTCTCAGCCA
CGGAAGGGGG	GGAGGGGGCA	GAAGAGAGGG	CGTCGACGGA	GTGAGCCGAT	GAGAGTCGGT
1150	1160	1170	1180	1190	1200
ACCCCCCTCA	CCACCCTTCT	CCCCACCCGC	CCCCCGGCC	CCGTCGGCCC	ABCGNTGNCA
TGGGGGGAGT	GGTGGGAAGA	GGGGTGGGCG	GGGGGGCGGG	GGCAGCCGGG	TCGCNACNGT
1210	1220	1230	1240	1250	1260
GNCCGAGTTT	GCAGAGAGGT	AACTCCCTTT	GGCTGCGAGC	GGGCGAGNCT	AGCTGCACAT
CNGGCTCAAA	CGTCTCTCCA	TTGAGGGAAA	CCGACGCTCG	CCCGCTCNBA	TCGACGTGTA
1270	1280	1290	1300	1310	1320
TGCAAAGAAG	GCTCTTAGGA	GCAGGCGACT	GGGGAGCGGC	TTCAGCACTG	CAGCCACGAC
ACGTTTCTTC	CGAGAATCCT	CGTCCGCTGA	CCCCTCGCCG	AAGTCGTGAC	GTGGGTGCTG
1330	1340	1350	1360	1370	1380
CNGCCTGSTT	AGGCTGCACG	CGGAGAGAAC	CCTCTGTTTT	CCCCCACTCT	CTCTCCACCT
GNCGGACCAA	TCCGACGTGC	GCCTCTCTTG	GGAGACAAAA	GGGGGTGAGA	GAGAGGTGGA
1390	1400	1410	1420	1430	1440
CCTCCTGCCCT	TCCCCACCCC	GAGTGCGGAG	CCAGAGATCA	AAAGATGAAA	AGGCAGTCAG
GGAGSACGGA	AGGGGTGGGG	CTCACGCCCTC	GGTCTCTAGT	TTTCTACTTT	TCCGTCACTC
1450	1460	1470	1480	1490	1500
GTCTTCAGTA	GCCAAAAAAC	AAAACAAACA	AAAACAAAAA	AGCCGAAATA	AAAGAAAAAG
CAGAGTCAAT	CGGTTTTTTG	TTTTGTTTTG	TTTTGTTTTT	TCGGCTTTAT	TTTCTTTTTT
1510	1520	1530	1540	1550	1560
ATAATACTC	AGTTCTTATT	TGCACCTACT	TCAGTGGACA	CTGAATTTGG	AAGBTGGAGG
TATTATTGAG	TCAAGAATAA	ACGTGGATGA	AGTCACCTGT	GACTTAAACC	TTCCACCTCC
1570	1580	1590	1600	1610	1620
ATTTTGTTTT	TTTCTTTTAA	GATCTGGGCA	TCTTTTGAAT	CTACCCTTCA	AGTATTAGAA
TAAAACAAAA	AAAGAAAAAT	CTAGACCCGT	AGAAAACTTA	GATGGGAAGT	TCATAATTCT
1630	1640	1650	1660	1670	1680
GACAGACTGT	GAGCCTAGCA	GGGCGAGATCT	TGTCCACCGT	GTGTCTTCTT	CTGCACGAGA
CTGTCTGACA	CTCGGATCGT	CCCCTCTAGA	ACAGGTGGCA	CACAGAAGAA	GACGTGCTCT
1690	1700	1710	1720	1730	1740
CTTTGAGGCT	GTCAGAGCGC	TTTTTGCGTG	GTTGCTCCCG	CAAGTTTCCT	TCTCTGGAGC
GAAACTCCGA	CAGTCTCGCG	AAAAACGCAC	CAACGAGGGC	GTTCAAAGGA	AGAGACCTCG
1750	1760	1770	1780	1790	1800
TTCCCGCAGG	TGGGCAGCTA	GCTGCAGCGA	CTACCCCATC	ATCACAGCCT	GTTGAACTCT
AAGGGCGTCC	ACCCGTCGAT	CGACGTCGCT	GATGGCGTAG	TAGTGTGCGA	CAACTTGAGA

[illegible]

1810	1820	1830	1840	1850	1860
TCTGAGCAAG	AGAAGGGGAG	GCGGGGTAAG	GGAAAGTAGGT	GGAAAGATTCA	GCCAAAGCTCA
AGACTCGTTT	TCTTCCCCTC	CGCCCCATTC	CCTTCATCCA	CCTTCTAAGT	CGGTTTCBAGT
1870	1880	1890	1900	1910	1920
AGGATGGAAAG	TGCAGTTAGG	GCTGGGAAGG	GTCTACCCCTC	GGCCGCGCTC	CAAGACCTAC
TCCTACCTTC	ACGTCAATCC	CGACCCTTCC	CAGATGGGAG	CCGGCGGCGAG	GTTCTGGGATG
1930	1940	1950	1960	1970	1980
CGAGGAGCTT	TCCAGAATCT	GTTCCAGAGC	GTGCGCGAAG	TGATCCAGAA	CCCGGGGCCCC
GCTCCTCGAA	AGGTCTTAGA	CAAGGTCTCG	CACGCGCTTC	ACTAAGTCTT	GGGCCCCGGGG
1990	2000	2010	2020	2030	2040
AGGCACCCAG	AGGCCGCGAG	CGCAGCACCT	CCCGGCGCCA	GTTTGCTGCT	GCTGCAAGCAG
TCCGTGGGTC	TCCGGCGCTC	GCCTCATGGA	GGGCGCGGGT	CAACGACGGA	CGACGTCGTC
2050	2060	2070	2080	2090	2100
CAGCAGCAGC	AGCAGCAGCA	GCAGCAGCAG	CAGCAGCAGC	AGCAGCAGCA	GCAGCAGCAG
GTGTCGTCG	TCGTCGTCGT	CGTCGTCGTC	GTGTCGTCG	TCGTCGTCGT	CGTCGTCGTC
2110	2120	2130	2140	2150	2160
CAGCAGCAAG	AGACTAGCCC	CAGGCAGCAG	CAGCAGCAGC	AGGGTGAGGA	TGGTTCTCCC
GTGTCGTTT	TCTGATCGGG	GTCCGTCGTC	GTGTCGTCG	TCCCACTCCT	ACCAAGAGGG
2170	2180	2190	2200	2210	2220
CAAGCCTCATC	GTAGAGGGCC	CACAGGCTAC	CTGGTCCCTGG	ATGAGGAACA	GCAACCTTCA
GTTCGGGTAG	CATCTCCGGG	GTGTCCGATG	GACCAAGACC	TACTCCTTGT	CGTTGGAGAT
2230	2240	2250	2260	2270	2280
CAGCCGCAAT	CGGCCCTGGA	GTGCCACCCC	GAGAGAGGTT	GGTCCCAGGA	GCCTGGAGCC
GTGCGGCTCA	GCCGGGACCT	CACGGTGGGG	CTCTCTCCAA	CGCAGGGTCT	CGGACCTCGG
2290	2300	2310	2320	2330	2340
GCTGTGCTCG	CCAGCAAGGG	GCTGCCGCGAG	CAGCTGCCAG	CACCTCCGGA	CGAGGATGAC
CGGCACGGGG	GGTGCTTCCC	CGACGGCGTC	GTGACGGGTC	GTGGAGGGCT	GCTCCTACTG
2350	2360	2370	2380	2390	2400
TCACTTCCCC	CATCCACGTT	GTCCCTGCTG	GGCCCCACTT	TCCCCGGCTT	AAGCAGCTGC
AGTCGACGGG	GTAGGTGCAA	CAGGGACGAC	CCGGGGTGAA	AGGGGGCGAA	TTCGTGACAG
2410	2420	2430	2440	2450	2460
TCCGCTGACC	TTAAAGACAT	CCTGAGCGAG	GCCAGCACCA	TGCAACTCCT	TCAGCAACAG
AGGCGACTGG	AATTTCTGTA	GGACTCGCTC	CGGTGCTGCT	ACGTTGAGGA	AGTCGTTGTC
2470	2480	2490	2500	2510	2520
CAGCAGGAAG	CAGTATCCGA	AGGCAGCAGC	ABCGGGAGAG	CGAGGGAGGC	CTCGGGGGCT
GTGTCCTTTC	GTCATAGGCT	TCCGTGCTCG	TCGCCCTCTC	GTCCTCTCCG	GAGCCCCCGA
2530	2540	2550	2560	2570	2580
CCCACTTCCCT	CCAAGGACAA	TACTTTAGGG	GGCACTTCGA	CCATTTCTGA	CAACGCCAAG
GGGTGAAGGA	GTTTCTCTGT	AATGAATCCC	CCGTGAAGCT	GGTAAGAGCT	GTTGCGGTTT
2590	2600	2610	2620	2630	2640
GAGTGTGTGA	AGGCAGTGTC	GGTGTCATG	GGCTGGGGTG	TGGAGGCGTT	GGAGCATCTG
CTCAACACAT	TCCGTCACAG	CCACAGGTAC	CCGGACCCAC	ACCTCCGCAA	CCTCGTAGAC
2650	2660	2670	2680	2690	2700
AGTCCAGGGG	AACAGCTTCG	GGGGGATTGC	ATGTACGCCC	CACTTTTGGG	AGTTCCACCC
TCAAGTCCCC	TTGTGGAAGC	CCCCCTAAGC	TACATGCGGG	GTGAAAACCC	TCAAGGTGGG

FIGURE 4 (page 6 of 6)

4510	4520	4530	4540	4550	4560
CAGTTCACTT	TTGACCTGCT	AATCAAGTCA	CACATGGTGA	GCGTGGACTT	TCCGGAAATG
GTCAAGTGAA	AACTGGACGA	TTAGTTCAGT	GTGTACCACT	CGCACCTGAA	AGGCCCTTAC
4570	4580	4590	4600	4610	4620
ATGGCAGAGA	TCATCTCTGT	GCAAGTGCCC	AAGATCCTTT	CTGGGAAAGT	CAAGCCCATC
TACCGTCTCT	AGTAGAGACA	CGTTCACGGG	TTCTAGGAAA	GACCCCTTCA	GTTCGGGTAG
4630	4640	4650	4660	4670	4680
TATTTCCACA	CCCAAGTGAAG	CATTGGAAAC	CCTATTTCCC	CACCCAGGCT	CATGCCCCCT
ATAAAGGIGT	GGGTCACTTC	GTAACCTTTG	GGATAAAGGG	GTGGGGTCGA	GTACGGGGGA
4690	4700	4710	4720	4730	4740
TTCAGATGTC	TTCTGCTGT	TATAACTCTG	CACTACTCCT	CTGCAGTGCC	TTGGGGAAAT
AAGTCTACAG	AAGACGGACA	ATATTGAGAC	GTGATGAGGA	GACGTACGGG	AACCCCTTAA
4750	4760	4770	4780	4790	4800
TCCTCTATTG	ATGTACAGTC	TGTATGAAC	ATGTTCTCTGA	ATTCTATTTG	CTGGGCTTTT
AGGAGATAAC	TACATGTGAG	ACAGTACTTG	TACAAGGACT	TAAGATAAAC	GACCCGAAAA
4810	4820	4830	4840	4850	4860
TTTTTCTCTT	TCTCTCCTTT	CTTTTTCTTC	TTCCCTCCCT	ATCTAACCCT	CCCATGGCAC
AAAAAGAGAA	AGAGAGGAAA	GAAAAAGAA	AAGGGAGGGG	TAGATTGGGA	GGGTACCGTG
4870	4880	4890	4900	4910	4920
CTTCAGACTT	TGCTTCCCAT	TGTGGCTCCT	ATCTGTGTTT	TGAATGGTGT	TGTATGCCCT
GAAGTCTGAA	ACGAAGGGTA	ACACCGAGGA	TAGACACAAA	ACTTACCACA	ACATACGGAA
4930	4940	4950	4960	4970	4980
TAAATCTGTG	ATGATCCTCA	TATGGCCGAG	TGTCAAGTTG	TGCTTGTTTA	CAGCACTACT
ATTTAGACAC	TACTAGGAGT	ATACCGGGTC	ACAGTTCAAC	ACGAACAAAT	GTGCTGATGA
4990	5000	5010	5020	5030	5040
CTGTGCCAGC	CACACAAACG	TTTACTTATC	TTATGCCACG	GGAAGTTTAG	AGAGCTAAGA
GACACGGTCG	GTGTGTTTGC	AAATGAATAG	AATACGGTGC	CCTTCAAATC	TCTCGATTCT
5050	5060	5070	5080		
TTATCTGGGG	AAATCAAAAC	AAAAACAAG	CAACCAAAAA	AAAAA	
AATAGACCCC	TTTAGTTTTG	TTTTTGTTC	TTTTTGTTC	TTTTT	

Variable	Mean	SD	Min	Max
Age	35.2	12.5	18	65
Gender	Male	10.5	0	20
Marital status	Married	15.2	0	25
Education	High school	12.5	0	20
Occupation	Unemployed	18.5	0	30
Income	Low	15.5	0	25
Health status	Good	12.5	0	20
Stress level	High	18.5	0	30
Life satisfaction	Low	15.5	0	25
Depression	High	18.5	0	30
Anxiety	High	18.5	0	30
Substance use	Low	15.5	0	25
Physical activity	Low	15.5	0	25
Social support	Low	15.5	0	25
Resilience	Low	15.5	0	25
Self-efficacy	Low	15.5	0	25
Optimism	Low	15.5	0	25
Gratitude	Low	15.5	0	25
Forgiveness	Low	15.5	0	25
Empathy	Low	15.5	0	25
Compassion	Low	15.5	0	25
Kindness	Low	15.5	0	25
Generosity	Low	15.5	0	25
Patience	Low	15.5	0	25
Humility	Low	15.5	0	25
Modesty	Low	15.5	0	25
Shyness	Low	15.5	0	25
Introversion	Low	15.5	0	25
Neuroticism	High	18.5	0	30
Extraversion	Low	15.5	0	25
Agreeableness	Low	15.5	0	25
Conscientiousness	Low	15.5	0	25
Openness	Low	15.5	0	25
Stability	Low	15.5	0	25
Emotion regulation	Low	15.5	0	25
Cognitive flexibility	Low	15.5	0	25
Problem solving	Low	15.5	0	25
Decision making	Low	15.5	0	25
Goal setting	Low	15.5	0	25
Time management	Low	15.5	0	25
Organization	Low	15.5	0	25
Planning	Low	15.5	0	25
Initiative	Low	15.5	0	25
Proactivity	Low	15.5	0	25
Resilience	Low	15.5	0	25
Self-efficacy	Low	15.5	0	25
Optimism	Low	15.5	0	25
Gratitude	Low	15.5	0	25
Forgiveness	Low	15.5	0	25
Empathy	Low	15.5	0	25
Compassion	Low	15.5	0	25
Kindness	Low	15.5	0	25
Generosity	Low	15.5	0	25
Patience	Low	15.5	0	25
Humility	Low	15.5	0	25
Modesty	Low	15.5	0	25
Shyness	Low	15.5	0	25
Introversion	Low	15.5	0	25
Neuroticism	High	18.5	0	30
Extraversion	Low	15.5	0	25
Agreeableness	Low	15.5	0	25
Conscientiousness	Low	15.5	0	25
Openness	Low	15.5	0	25
Stability	Low	15.5	0	25
Emotion regulation	Low	15.5	0	25
Cognitive flexibility	Low	15.5	0	25
Problem solving	Low	15.5	0	25
Decision making	Low	15.5	0	25
Goal setting	Low	15.5	0	25
Time management	Low	15.5	0	25
Organization	Low	15.5	0	25
Planning	Low	15.5	0	25
Initiative	Low	15.5	0	25
Proactivity	Low	15.5	0	25
Resilience	Low	15.5	0	25
Self-efficacy	Low	15.5	0	25
Optimism	Low	15.5	0	25
Gratitude	Low	15.5	0	25
Forgiveness	Low	15.5	0	25
Empathy	Low	15.5	0	25
Compassion	Low	15.5	0	25
Kindness	Low	15.5	0	25
Generosity	Low	15.5	0	25
Patience	Low	15.5	0	25
Humility	Low	15.5	0	25
Modesty				

10	20	30	40	50	60
GAGCTCTGGA	CAAAATTGAG	CGCCTATGTG	TACATGGCAA	GTGTTTTTAG	TGTTTGTGTG
70	80	90	100	110	120
TTTACCTGCT	TGTCTGGGTG	ATTTTGCCTT	TGAGAGTCTG	GATGAGAAAT	GCATGGTTAA
130	140	150	160	170	180
AGGCAATTCC	AGACAGGAAG	AAAGGCAGAG	AAGAGGGTAG	AAATGACCTC	TGATTCTTGG
190	200	210	220	230	240
GGCTGAGGGT	TCCTAGAGCA	AATGGCACAA	TGCCACGAGG	CCCGATCTAT	CCCTATGACG
250	260	270	280	290	300
GAACTCTAAG	GTTTCAGCAT	CAGCTATCTG	CTGGCTTGGT	CAGTGGCTTG	CCTCCTCAGT
310	320	330	340	350	360
TTGTAGGAGA	CTCTCCCACT	CTCCCATCTG	CGCGCTCTTA	TCAGTCTCTA	AAAGAACCCN
370	380	390	400	410	420
TGGCNAGCCA	GGAGCNAGGT	ATTGNTATCG	TCCTTTTCNT	CCTCCTNGCC	TCACCTNGTT
430	440	450	460	470	480
GNTTTTTAGA	TTGGNCTTNG	NAACCAAATT	TGTATGCTGG	CCTCCAGGAA	ATCTGGAGCC
490	500	510	520	530	540
TGGCGCCTAA	ACCTTGGTTT	AGGAAAGCAG	GAGCTATTCA	GGAGGCAGGG	TCCTCCAGGG
550	560	570	580	590	600
CTAGAGCTAG	CCTCTCCTGC	CCTCGCCAC	GTGCGCCAGC	ACTTGTCTCT	CCAAAGCNAC
610	620	630	640	650	660
TAGSCAGGCG	TTAGCGCGCG	GTGAGGGGAG	GGGAGAAAAG	GAAAGGGGAG	GGGAGGGAAA
670	680	690	700	710	720
AGGAGGTGGG	AAGGCAAGGA	GGCCGGCCNG	GTGGGGGGCG	GACCCGACTC	GCANNAACTG
730	740	750	760	770	780
TTGCATTTGC	TCTCCACCTC	CCAGCGCCCC	CTCCGAGATC	CCGGGGAGCC	AGCTTGCTGG
790	800	810	820	830	840
GAGAGCGGGA	ACGGTCCGGA	GCAAGCCCAG	AGGCAGAGGA	GGCGACAGAG	GGAAAAAGGG
850	860	870	880	890	900
CCCNAGCTAG	CCGCTCCAGT	GCTGTACAGN	AGCCGAAGGA	CGCACCACGC	CAGCCCCAGC
910	920	930	940	950	960
CCGGCTCCAG	CGACAGCNAA	CGCCTCTTGC	ANGCGTTTGA	AGCCGCCGCC	CGGAGCTGCC
970	980	990	1000	1010	1020
CTTTCCTCTT	CGGTGAAGTT	TTTAAAAGCT	GCTAAAGACT	CGGAGGAAGC	AAGGAAAGTG
1030	1040	1050	1060	1070	1080
CCTGGTAGGA	CTGACGGCTG	CCTTTGTCTT	CCTCCTCTCC	ACCCCGCCTC	CCCCACCCTT
1090	1100	1110	1120	1130	1140
GCCTTCCCCC	CCTCCCCCGT	CTTCTCTCCC	GCAGCTGCCT	CAGTCGGCTA	CTCTCAGCCA
1150	1160	1170	1180	1190	1200
ACCCCCCTCA	CCACCTTTCT	CCCCACCCGC	CCCCCGCCCC	CCGTGGGCCC	AGCGNTGNCA
1210	1220	1230	1240	1250	1260
GNCCGAGTTT	GCAGAGAGGT	AAGTCCCTTT	AGCTGCGAGC	GGGCGAGNCT	AGCTGACATG

[illegible]

1270	1280	1290	1300	1310	1320
TGC4AAGAAG	GCTCTTAGGA	GCAGGCGACT	GGGGAGCGGC	TTCAGCACTG	CAGCCACGAC
1330	1340	1350	1360	1370	1380
CNGCCTGGTT	AGGCTGCACG	CGGAGAGAAC	CCTCTGTTTT	CCCCCACTCT	CTCTCCACCT
1390	1400	1410	1420	1430	1440
CCTCCTGCCT	TCCCCACCCC	GAGTGCGGAG	CCAGAGATCA	AAAGATGAAA	AGGCAGTCAG
1450	1460	1470	1480	1490	1500
GTETTCAGTA	GCCAAAAAAC	AAAACAAACA	AAAACAAAAA	AGCCGAAATA	AAAGAAAAAG
1510	1520	1530	1540	1550	1560
ATAATAALTC	AGTTCTTATT	TGCACCTACT	TCAGTGGACA	CTGAATTTGG	AAGGTGGAGG
1570	1580	1590	1600	1610	1620
ATTTTGTITT	TTTCTTTTAA	GATCTGGGCA	TCTTTTGAAT	CTACCCCTTCA	AGTATTAASA
1630	1640	1650	1660	1670	1680
GACAGACTGT	GAGCCTAGCA	GGGCAGATCT	TGTCCACCGT	GTGTCTTCTT	CTGCACGAGA
1690	1700	1710	1720	1730	1740
CTTTGAGGCT	GTCAGAGCGC	TTTTTGCGTG	GTTGCTCCCG	CAAGTTTCCT	TCTCTGGAGC
1750	1760	1770	1780	1790	1800
TTCCCGCAGG	TGGGCAGCTA	GCTGCAGCGA	CTACCGCATC	ATCACAGCCT	GTTGACTCT
1810	1820	1830	1840	1850	1860
TCTGAGCAGG	AGAAAGGGAG	GCGGGGTAAG	GGAGGTAGGT	GGAGGATTCA	GCCAAGCTCA
1890					
GGG	ATG	GAA	GTG	CAG	TTA
1st	Glu	Val	Gln	Leu	Gly
1950					
GGG	CTG	GGA	AGG	GTG	TAC
1st	Glu	Val	Gln	Leu	Gly
1980					
GGG	CTG	GGA	AGG	GTG	TAC
1st	Glu	Val	Gln	Leu	Gly
2010					
GGG	CTG	GGA	AGG	GTG	TAC
1st	Glu	Val	Gln	Leu	Gly
2040					
GGG	CTG	GGA	AGG	GTG	TAC
1st	Glu	Val	Gln	Leu	Gly
2070					
GGG	CTG	GGA	AGG	GTG	TAC
1st	Glu	Val	Gln	Leu	Gly
2100					
GGG	CTG	GGA	AGG	GTG	TAC
1st	Glu	Val	Gln	Leu	Gly
2130					
GGG	CTG	GGA	AGG	GTG	TAC
1st	Glu	Val	Gln	Leu	Gly
2160					
GGG	CTG	GGA	AGG	GTG	TAC
1st	Glu	Val	Gln	Leu	Gly
2190					
GGG	CTG	GGA	AGG	GTG	TAC
1st	Glu	Val	Gln	Leu	Gly
2220					
GGG	CTG	GGA	AGG	GTG	TAC
1st	Glu	Val	Gln	Leu	Gly
2250					
GGG	CTG	GGA	AGG	GTG	TAC
1st	Glu	Val	Gln	Leu	Gly
2280					
GGG	CTG	GGA	AGG	GTG	TAC
1st	Glu	Val	Gln	Leu	Gly
2310					
GGG	CTG	GGA	AGG	GTG	TAC
1st	Glu	Val	Gln	Leu	Gly
2340					
GGG	CTG	GGA	AGG	GTG	TAC
1st	Glu	Val	Gln	Leu	Gly

FIGURE 5 (page 4 of 5)

[illegible]

3330 3360
GCT GTA GCC CCC TAC GGC TAC ACT CGG CCC CCT CAG GGG CTG GCG GGC CAG GAA AGC GAC
Ala Val Ala Pro Tyr Gly Tyr Thr Arg Pro Pro Gln Gly Leu Ala Gly Gln Glu Ser Asp

3390 3420
TTC ACC GCA CCT GAT GTG TGG TAC CCG GGC GGC ATG GTG AGC ABA GTG CCC TAT CCC AGT
Phe Thr Ala Pro Asp Val Trp Tyr Pro Gly Gly Met Val Ser Arg Val Pro Tyr Pro Ser

3450 3480
CCC ACT TGT GTC AAA AGC GAA ATG GGC CCC TGG ATG GAT AGC TAC TCC CGG GAA CCT TAC
Pro Thr Cys Val Lys Ser Glu Met Gly Pro Trp Met Asp Ser Tyr Ser Arg Glu Pro Tyr

3510 3540

GGG GAC ATG CGT TTG GAG ACT GCC AGG GAC CAT GTT TTG CCC ATT GAC TAT TAC TTT CCA
Gly Asp Met Arg Leu Glu Thr Ala Arg Asp His Val Leu Pro Ile Asp Tyr Tyr Phe Pro

3570 3600
CCC CAG AAG ACC TGC CTG ATC TGT GGA GAT GAA GCT TCT GGG TGT CAC TAT GGA GCT CTC
Pro Gln Lys Thr Cys Leu Ile Cys Gly Asp Glu Ala Ser Gly Cys His Tyr Gly Ala Leu

ACA TGT GGA AGC TGC AAG GTC TTC TTC AAA AGA GCC GCT GAA GGG AAA CAG AAG TAC CTG
Thr Cys Gly Ser Cys Lys Val Phe Phe Lys Arg Ala Ala Glu Gly Lys Gln Lys Tyr Leu

3690 3720
TGC GCC AGC AGA AAT GAT TGC ACT ATT GAT AAA TTC CGA AGG AAA AAT TGT CCA TCT TGT
Cys Ala Ser Arg Asn Asp Cys Thr Ile Asp Lys Phe Arg Arg Lys Asn Cys Pro Ser Cys

CGT CTT CGG AAA TGT TAT GAA GCA GGG ATG ACT CTG GGA GCC CCG AAG CTG AAG AAA CTT
Arg Leu Arg Lys Cys Tyr Glu Ala Gly Met Thr Leu Gly Ala Arg Lys Leu Lys Lys Leu

3810 3840
GGT AAT CTG AAA CTA CAG GAG GAA GGA GAG GCT TCC AGC ACC ACC AGC CCC ACT GAG GAG
Gly Asn Leu Lys Leu Gln Glu Glu Gly Glu Ala Ser Ser Thr Thr Ser Pro Thr Glu Glu

3870 3900

ACA ACC CAG AAG CTG ACA GTG TCA CAC ATT GAA GGC TAT GAA TGT CAG CCC ATC TTT CTG
Thr Thr Gln Lys Leu Thr Val Ser His Ile Glu Gly Tyr Glu Cys Gln Pro Ile Phe Leu

3930 3960
AAT GTC CTG GAA GCC ATT GAG CCA GGT GTA GTG TGT GCT GGA CAC GAC AAC AAC CAG CCC
Asn Val Leu Glu Ala Ile Glu Pro Gly Val Val Cys Ala Gly His Asp Asn Asn Gln Pro

3990 4020

GAC TCC TTT GCA GCC TTG CTC TCT AGC CTC AAT GAA CTG GGA GAG AGA CAG CTT GTA CAC
Asp Ser Phe Ala Ala Leu Leu Ser Ser Leu Asn Glu Leu Gly Glu Arg Gln Leu Val His

4050 4080

GTG GTC AAG TGG GCC AAG GGC TTG CCT GGC TTC CGC AAC TTA CAC GTG GAC GAC CAG ATG
Val Val Lys Trp Ala Lys Gly Leu Pro Gly Phe Arg Asn Leu His Val Asp Asp Gln Met

4110 4140
GCT GTC ATT CAG TAC TCC TGG ATG GGG CTC ATG GTG TTT GCC ATG GGC TGG CGA TCC TTC
Ala Val Ile Gln Tyr Ser Trp Met Gly Leu Met Val Phe Ala Met Gly Trp Arg Ser Phe

FIGURE 5 (page 5 of 5)

4170 4200
ACC AAT GTC AAC TCC AGG ATG CTC TAC TTC GCC CCT GAT CTG GTT TTC AAT GAG TAC CGC
Thr Asn Val Asn Ser Arg Met Leu Tyr Phe Ala Pro Asp Leu Val Phe Asn Glu Tyr Arg

4230 4260
ATG CAC AAG TCC CGG ATG TAC AGC CAG TGT GTC CGA ATG AAG CAC CTC TCT CAA GAG TTT
Met His Lys Ser Arg Met Tyr Ser Gln Cys Val Arg Met Arg His Leu Ser Gln Glu Phe

4290 4320
GGA TGG CTC CAA ATC ACC CCC CAG GAA TTC CTG TGC ATG AAA GCA CTG CTA CTC TTC AGC
Gly Trp Leu Gln Ile Thr Pro Gln Glu Phe Leu Cys Met Lys Ala Leu Leu Leu Phe Ser

4350 4380
ATT ATT CCA GTG GAT GGG CTG AAA AAT CAA AAA TTC TTT GAT GAA CTT CGA ATG AAC TAC
Ile Ile Pro Val Asp Gly Leu Lys Asn Gln Lys Phe Phe Asp Glu Leu Arg Met Asn Tyr

4410 4440
ATC AAG GAA CTC GAT CGT ATC ATT GCA TGC AAA AGA AAA AAT CCC ACA TCC TGC TCA AGA
Ile Lys Glu Leu Asp Arg Ile Ile Ala Cys Lys Arg Lys Asn Pro Thr Ser Cys Ser Arg

4470 4500
CGC TTC TAC CAG CTC ACC AAG CTC CTG GAC TCC GTG CAG CCT ATT GCG AGA GAG CTG CAT
Arg Phe Tyr Gln Leu Thr Lys Leu Leu Asp Ser Val Gln Pro Ile Ala Arg Glu Leu His

4530 4560
CAG TTC ACT TTT GAC CTG CTA ATC AAG TCA CAC ATG GTG AGC GTG GAC TTT CCG GAA ATG
Gln Phe Thr Phe Asp Leu Leu Ile Lys Ser His Met Val Ser Val Asp Phe Pro Glu Met

4590 4620
ATG GCA GAG ATC ATC TCT GTG CAA GTG CCC AAG ATC CTT TCT GGG AAA GTC AAG CCC ATC
Met Ala Glu Ile Ile Ser Val Gln Val Pro Lys Ile Leu Ser Gly Lys Val Lys Pro Ile

4650 4680
TAT TTC CAC ACC CAG TGA AGC ATT GGA AAC CCT ATT TCC CCA CCC CAG CTC ATG CCC CCT
Tyr Phe His Thr Gln End

4690 4700 4710 4720 4730 4740
TTCAGATGTC TTCTGCCTGT TATAACTCTG CACTACTCCT CTGCAGTGCC TTGGGGGAATT

4750 4760 4770 4780 4790 4800
TCCTCTATTG ATGTACAGTC TGTCATGAAC ATGTTCTCTGA ATTCTATTTG CTGGGCTTTT

4810 4820 4830 4840 4850 4860
TTTTTCTCTT TCTCTCCTTT CTTTTTCTTC TTCCCTCCCT ATCTAACCCT CCCATGGCAC

4870 4880 4890 4900 4910 4920
CTTCAGACTT TGCTTCCCAT TGTGGCTCCT ATCTGTGTTT TGAATGGTGT TGTATGCCTT

4930 4940 4950 4960 4970 4980
TAAATCTGTG ATGATCCTCA TATGGCCAG TGTCAAGTTG TGCTTGTTTA CAGCACTACT

4990 5000 5010 5020 5030 5040
CTGTGCCAGC CACACAAACG TTTACTTATC TTATGCCACG GGAAGTTTAA AGAGCTAAGA

5050 5060 5070 5080
TTATCTGEGG AAATCAAAAC AAAAAACAAG CAACAAAAA AAAAA

FIGURE 6 (page 1075)

[illegible][illegible]

930 935 940 945 950 955 960

AAATAGCTAGGACAGATCCCTCTCTATAGCGCTGCTTCTCTTTACAGGAGACTTTTGAAGCTA

970 980 990 1000 1010 1020

TTTGAGGCTCTCCCGCCCCCTCCCTGCGAAGTTTTCTTCTCCTGGAAGCTTCCCGCAGGTTGGGCA

1030 1040 1050 1060 1070 1080

CCCTAGCTGCAGATACCTACATCATCACTCAGTAGGACTCTTTCAGAGCAAGAGAGACGAGGAGG

1090 1100 1110 1120 1130 1140

CAGGATTAAGCGAATTCTGCTGCGAAGCTAGAGAGCAAGCTTAAGGAGATGGAGGTTGCAATTAAGG

MetGluValGlnLeuGly

1150 1160 1170 1180 1190 1200

CTGGGAAAGGCTCTACCCACAGCTCCCCGTCCAAAGACCTATCGAGGAGCGTTCCAGAAATCTG

LeuGlyArgValTyrProArgProProSerLysThrTyrArgGlyAlaPheGlnAsnLeu

1210 1220 1230 1240 1250 1260

TTCCAGAGCGCTGCGCGAAGCGATCCAGAAACCCCGGGCCCCAGGCAACCCCTGAGGCGCGCTAGC

PheGlnSerValArgGluAlaIleGlnAsnProGlyProArgHisProGluAlaAlaSer

1270 1280 1290 1300 1310 1320

ATAGCACTCTCCCGGTGCTCTGTTTTACAGCAAGCGGAGGAGAGACTAGCTCCCTCGGGGCGCGCGG

IleAlaProProGlyAlaCysLeuGlnGlnArgGlnGluThrSerProArgArgArgArg

1330 1340 1350 1360 1370 1380

GGGCAAGCAAGCCCTGAGGATGCGCTCTCTCAAGCCCCACATCAGAAAGCAACCCACAGGCTAC

ArgGlnGlnHisProGluAspGlySerProGlnAlaHisIleArgGlyThrThrGlyTyr

1390 1400 1410 1420 1430 1440

CTGGGCTCTGAGGAGGGAACAGCAGCCTTCAGAGCAGCAGTCAAGCTTCCGAGGGCCACCCCT

LeuAlaLeuGluGluGluGlnGlnProSerGlnGlnGlnSerAlaSerGluGlyHisPro

1450 1460 1470 1480 1490 1500

GAGAGCGGCTGCTTCCCGAAGCCTGAGAGCTGCCACGGCTCTTGGCAAGGGGCTGCCGCAAG

GluSerGlyCysLeuProGluProGlyAlaAlaThrAlaProGlyLysGlyLeuProGln

1510 1520 1530 1540 1550 1560

CAGCCAGCAGCTCCTTCAGATCAGGATGACTCAGCTGCCCAATCCAGGTTGTCCTCTACTG

GlnProProAlaProProAspGlnArgAspSerAlaAlaProSerThrLeuSerLeuLeu

1570 1580 1590 1600 1610 1620

GGCCCGCACTTTCACAGGCTTAAGCAAGCTGCTCCGCAAGACATTAAAGACATCCTGAGCGAG

GlyProThrPheProGlnLeuSerSerLysSerAlaAspIleLysAspIleLeuSerGlu

1630 1640 1650 1660 1670 1680

CCGCGTACCTTGCAGCTTCTTCAGCAAGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG

AlaGlyThrThrGlnLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln

1690 1700 1710 1720 1730 1740

CAGCAAGCAGCGAGCAAGCAAGCAGCAGCAGGTAATATCCGAGAGGCAAGCAAGCAAGCAAGCAAG

Gln

1750 1760 1770 1780 1790 1800

GCAAGGAGAGGCGCATGGGCGCTCCCTCTTCCCTCCAGGATAGTTTCCTAGGGGCGCAATTCCG

AlaArgGlnAlaThrGlyAlaProIleSerSerSerLysAspSerTyrLeuGlyGlyAsnSer

1810	1820	1830	1840	1850	1860
ACCATCTCTGACAGTGCACAAGGAGTTGTGTAAAGCACTGCTCTGTGTCCATGGGGTTGGGT					
ThrIleSerAspSerAlaLysGluLeuCysLysAlaValSerValSerMetGlnLeuGly					
1870	1880	1890	1900	1910	1920
GTGGAAAGTACTGGAAACATCTGAGTCCAGGGGAGCAGCTTCGGGGGAGTGCATGTACGCG					
ValGluAlaLeuGluHisLeuSerProGluGluGlnLeuArgGlyAspCysMetTyrAla					
1930	1940	1950	1960	1970	1980
TCCCTCCTGGTGGTCCACCGGCCGTGCGTCCCACTCCTTTGTGCGCCTCTGGCCGAATGC					
SerLeuLeuGlyGlyProProAlaValArgProThrProCysAlaProLeuAlaGluCys					
1990	2000	2010	2020	2030	2040
CAAGGTCTTTCCCTGGACGAGAGCCCCGGGCAAAAGGCACTGAAGAGACTGCTGGATATTCC					
LysGlnLeuSerLeuAspGluGlyProGlyLysGlnThrGluGluThrAlaGluLysSer					
2050	2060	2070	2080	2090	2100
TCTTTCAAGGGAGGTTACGCCAAGGCTTGGAAAGGTGAGAGTCTGCGCTGCTCTGGCAGC					
SerPheLysGluGlyTyrAlaLysGluLeuGluGlyGluSerLeuGlyCysSerGlySer					
2110	2120	2130	2140	2150	2160
AGTGAAGCAGGTAGCTCTGGGACACTTGAGATCCCGTCTCTCACTGTCTCTGTATAAGTCT					
SerGluAlaGlySerSerGlyThrLeuGluIleProSerSerLeuSerLeuTyrLysSer					
2170	2180	2190	2200	2210	2220
TCAGCACTAGACGAGGCAGCAGCATACCAAGATCGCGCACTACTACACCTTTCCCTCTGCT					
GluAlaValAspGluAlaAlaAlaTyrGlnAsnProAspTyrTyrAsnPheProLeuAla					
2230	2240	2250	2260	2270	2280
CTGTCTGGGCGCCGCAACCCCGCGCCCGCTACCCATCCACAGGCGCGCATCAAGCTGGAG					
LeuSerGlyPheProHisProProProProThrHisProHisAlaArgIleLysLeuGlu					
2290	2300	2310	2320	2330	2340
AACCTCTCGGACTACGGCAGCGCCTGGGCTGCGGCGGGCAGCGCAATGCCGCTATGGGGAC					
AsnProSerAspTyrGlySerAlaTrpAlaAlaAlaAlaAlaGlnCysArgTyrGlyAsp					
2350	2360	2370	2380	2390	2400
TTGGCTAGCCTACATGGAGGGAGTGTAGCGCGGACCCAGCACTGGATCGCGCCCAAGCACC					
LeuAlaSerLeuHisGlyGlySerValAlaGlyProSerThrGlySerProProAlaThr					
2410	2420	2430	2440	2450	2460
GCTCTCTCTTCTCTGGCATACTCTCTTCACAGCTCAAGGAAGGCCAATTATATGGGCCAGGA					
AlaSerSerSerSerTrpHisThrLeuPheThrAlaGluGluGluGlnLeuTyrGlyProGly					
2470	2480	2490	2500	2510	2520
CGCTGAGGGGACCTGCAGTAGTCCCAAGGCGATGCTGGGCTGTAGCCCCCTATGGCTACACT					
GlyGluGlyCysSerSerSerProSerAspAlaGlyProValAlaProTyrGlyTyrThr					
2530	2540	2550	2560	2570	2580
CGCTCCCGCTAGGGGCTGCCAAGGCCAGGAGGGTCACTTCTCTGCTCTGGATTTGTGATAT					
ArgProProGluGlyLeuAlaSerAlaGluGlyAspPheSerAlaSerGluValTrpTyr					
2590	2600	2610	2620	2630	2640
CTGGTGGAGTGTGTGACAGAGTCCCTATCCAGTCCCAAGTTGTGTTAAGAGTGAATG					
ProGlyProLysValAlaAsnArgValProTyrProSerProSerCysValLysSerGluMet					
2650	2660	2670	2680	2690	2700
GGACCTTCTGATGGAGAACTACTCCGGACCTTATGGGACATGCGTTTGACCACTACCAAGG					
GlyProTrpMetGluAsnTyrSerGlyProTyrGlyAsnMetArgLeuAspSerThrArg					

2710	2720	2730	2740	2750	2760
GACCCACCGTTTACCCCATCGACTATTACTTCCCACCCACAGAGACCTTCCTGTCCTGTGGA					
AsnHisValLeuProIleAspTyrTyrPheProProGlnLysThrCysLeuIleCysGly					
2770	2780	2790	2800	2810	2820
GATGACGCTTCTGGTTGTCACTACGGAGCTCTCACTTGTGTGCAGCTGCCAAGGTCTTCTTC					
AspGluAlaSerGlyCysHisTyrGlyAlaLeuThrCysGlySerCysLysValPhePhe					
2830	2840	2850	2860	2870	2880
AAAGAGAGCTTCCGGGAAGGGAAACAGAAATATCTATGTGCCAGCAGAAATGATTGCACCAT					
LysArgAlaIleGluGlyLysGlnLysTyrLeuCysAlaSerArgAsnAspCysThrIle					
2890	2900	2910	2920	2930	2940
GATAAATTTTCGGAGAGAAAAATTGTCCATCCTGTCTGTCTCCGGAATGTATATGAAGCAGGG					
AspLysPheArgArgLysAsnCysProSerCysArgLeuArgLysCysTyrGluAlaGly					
2950	2960	2970	2980	2990	3000
ATGACTCTCGGAGCTTCGTAAGCTGAAGAAACTTTGGAATCTCAAACACAGGAAAGAGGA					
MetThrLeuGlyAlaArgLysLeuLysLysLeuGlyAsnLeuLysLeuGlnGluGluGly					
3010	3020	3030	3040	3050	3060
GAAAGCTTCAGTGTCTGTAAGCCCACTGAGGACCCATCCACAGAGATGACTGTATCACAC					
GluAsnSerSerAlaGlySerProThrGluAspProSerGlnLysMetThrValSerHis					
3070	3080	3090	3100	3110	3120
ATTGAAGCCATATGAATGTCAACTATCTTTTCTTAATGTCTTGGAAGCCATTTCGCCAGGA					
IleGluGlyTyrGluCysGlnProIlePheLeuAsnValLeuGluAlaIleGluProGly					
3130	3140	3150	3160	3170	3180
GTGGTGTGTGCGGGACATGACACCAACGAGCCTGATTCTCTTGTCTGCTTGTATCTAGT					
ValValCysAlaGlyHisAspAsnAsnGlnProAspSerPheAlaAlaLeuLeuSerSer					
3190	3200	3210	3220	3230	3240
CTCAACGAGCTTTGGCGAGAGACAGCTTGTACATGTGTGTCAGGTGGGCCCAAGGCTTTGCCT					
LeuAsnGluLeuGlyGluProGlnLeuValHisValValLysTrpAlaLysAlaLeuPro					
3250	3260	3270	3280	3290	3300
GGCTTCCGCAACTTGCATGTGGATGACCAAGATGGCAGTCATTCAGTATTCTCTCATGGGA					
GlyPheArgAsnLeuHisValAspAspGlnMetAlaValIleGlnTyrSerTrpMetGly					
3310	3320	3330	3340	3350	3360
CTGATGCTATTTTGCCATGGGTTGGCGGTCTTCACTAATGTCAACTCTAGGATGCTCTAC					
LeuMetValPheAlaMetGlyTrpArgSerPheThrAsnValAspSerArgMetLeuTyr					
3370	3380	3390	3400	3410	3420
TTTGCACCTGACCTGGTTTTTCAATGAGTATCGCATGGCACTAGTCTCGAATGACAGCCAG					
PheLeuProAspLeuValPhePheGlyTyrArgMetHisLysSerArgMetTyrSerGln					
3430	3440	3450	3460	3470	3480
TCTGTCTGCTGAGGACCTTCTCTCAAGCTTTTGAATGCTCCAGGATACCCCTCCAGGAA					
CysHisArgHisArgHisLeuSerGlnGluPheGlyTyrLeuGlnIleThrProGlnGlu					
3490	3500	3510	3520	3530	3540
TTCTTGTGCTGAAAGCACTCTACTCTTCAGCATTAATTCAGTGGATGGGCTGAAAGAAAT					
PheLeuLysMetLysAlaLeuLeuLeuPheSerIleIleProValAspGlyLeuLysAsn					
3550	3560	3570	3580	3590	3600
CAGAGATCTTTTGATGAATTCGAATGAAGTACATCAAGGAACTTGAATGCACTATTGCA					
GlnLysPhePheAspGluLeuArgMetAsnTyrIleLysGluLeuAspAspIleIleAla					

[illegible]

FIGURE 7



Figure 7. Frozen section of rat ventral prostate stained with antibodies (AR-52-3-p) to the AR peptide $\text{NH}_2\text{-Asp-His-Val-Leu-Pro-Ile-Asp-Tyr-Tyr-Phe-Pro-Pro-Gln-Lys-Thr}$ in a dilution of 1 to 3000 using the avidin-biotin peroxidase technique. Androgen receptor is indicated by brown staining of nuclei in epithelial cells. Immuno-staining was performed as previously described (60).

FIGURE 8

Restriction Fragment Length Polymorphism in the Human Androgen Receptor Gene

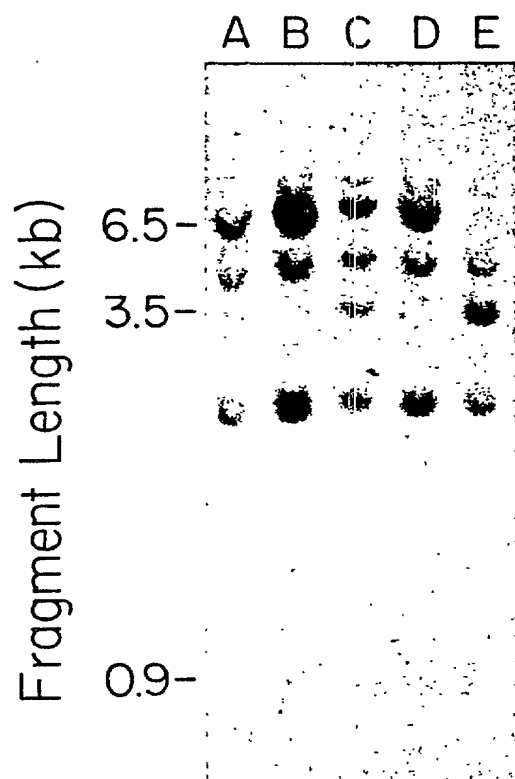


FIGURE 9

**Southern Blot Analysis of
Complete Androgen Insensitivity
Syndrome Patients**

